SECTORAL PRICES AND PRICE-SETTING

by

CHAD THOMAS FULTON

A DISSERTATION

Presented to the Department of Economics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2016

DISSERTATION APPROVAL PAGE

Student: Chad Thomas Fulton

Title: Sectoral Prices and Price-setting

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

Bruce McGough	Co-Chair
George Evans	Co-Chair
Jeremy Piger	Core Member
Stephen McKeon	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2016

DISSERTATION ABSTRACT

Chad Thomas Fulton

Doctor of Philosophy

Department of Economics

June 2016

Title: Sectoral Prices and Price-setting

This dissertation explores the price-setting behavior of firms both theoretically and empirically. The first portion constructs a theoretical model of price-setting in which firms are rationally inattentive: they cannot perfectly attend to all sources of uncertainty. By accommodating multiple sources of uncertainty within the model, it is possible to reasonably calibrate key parameters of the model. This bolsters the case for rational inattention as a microfounded alternative to ad-hoc mechanisms in order to generate price-stickiness and it not only allows for multiple sectors but demonstrates why their introduction is important.

The second portion contributes to the empirical literature exploring disaggregated price series. Taking into account the lessons from the theoretical model, a combination of dynamic factor and unobserved component models are applied to explicitly model heterogenous dynamic processes for sectoral prices. The key finding is that models with enforced homogenous dynamics are outperformed under a variety of criteria. More importantly, models with enforced homogenous dynamics can generate erroneous conclusions with respect to the speed of price responses to aggregate and idiosyncratic shocks.

A large body of recent empirical work on price-setting, including the empirical exercise described above, estimates a dynamic factor model using a relatively simple and partially non-parametric method. This method is valid in large samples, but alternative parametric methods exist that may be more efficient in small samples. The final portion of this dissertation compares methods for the estimation of dynamic factor models, including non-parametric, classical, and Bayesian techniques. The results of a Monte Carlo experiment validate the use of the partially non-parametric method, but find that the Bayesian approach may provide weakly superior results.

CURRICULUM VITAE

NAME OF AUTHOR: Chad Thomas Fulton

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR

DEGREES AWARDED:

Doctor of Philosophy, Economics, 2016, University of Oregon
Master of Science, Economics, 2012, University of Oregon
Bachelor of Arts, Economics, 2009, University of Oregon

AREAS OF SPECIAL INTEREST:

Macroeconomics and time series econometrics

GRANTS, AWARDS, AND HONORS:

Best Ph.D. Field Paper, 2015

Kleinsorge Summer Research Fellowship, 2015

Kleinsorge First Year Fellowship, 2011

ACKNOWLEDGEMENTS

I thank Professors George Evans, Bruce McGough, and Jeremy Piger for their superb instruction, advice, intuition, and support.

For Brittany and Basil.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Recently empirical research has emphasized that theoretical models of price-setting must distinguish between the effects of aggregate and sector-level shocks, and moreover that they must support heterogeneous behavior across sectors. The second chapter of this dissertation develops a model that can deliver these features by extending the rational inattention price-setting approach pioneered by Makowiak and Wiederholt (2009) to a multisector setting. Our analytic solution to a special case of the rational inattention problem allows us to detail attention allocation mechanisms and explore implications. More generally, we find that the multisector setting preserves the desirable characteristic that firms respond differently to different types of shocks, allows for heterogeneous responses, and reduces the need for extreme calibration of key parameters.

The third chapter of this dissertation examines the reduced form dynamics of disaggregated inflation series. We present a unified framework for understanding these dynamics under the heading of unobserved components models. We consider a wide range of candidate models to establish stylized facts, finding that inflation series have sufficiently heterogeneous dynamics that this wide range of models is necessary to adequately describe them. Equipped with the estimated dynamics, we find no evidence of pervasive measurement error that substantially distorts estimated stochastic characteristics. Finally, we buttress the proposition that sector-level prices respond quickly to idiosyncratic shocks while responding slowly to aggregate shocks, a point that was a crucial motivating factor for the preceding chapter but had been recently called into question.

The fourth chapter explores the finite-sample performance of dynamic factor models, considering two estimators not previously examined in this context - maximum likelihood using quasi-Netwon optimization and Bayesian Gibbs sampling - and constructing a more general Monte Carlo setup than has previously been used so as to explore performance under a wider range of empirically relevant data generating processes. The Monte Carlo analysis suggests the Gibbs sampling procedure is the weakly superior method but confirms previous studies that find generally good performance across estimators as the sample size becomes large; this holds even in the general data generating processes we consider here. We find that a simple two-step estimator used in the third chapter, above, is sufficient to adequately estimate factors are provide forecasts, providing a justification for using it rather than a more complicated estimators. In addition to evaluting estimators based on summary statistics, we look at the distribution across Monte Carlo replications to examine poor performance of maximum likelihood estimation in estimating factors with complex dynamics. If factor estimates are the objects of interest, our results support the common practice of removing outliers, but we find it to be detrimental to forecasting.

CHAPTER II

OPTIMAL PRICES IN MULTISECTOR MODELS UNDER RATIONAL INATTENTION

Introduction

A large branch of macroeconomic literature is concerned with the apparent non-neutrality of monetary policy in the short-run, addressing the question of why *nominal* changes have *real* effects.[1] This literature stretches back to Keynes who suggested wage and price stickiness as a mechanism by which an economy might fail to fully and immediately adjust to nominal changes - and, thus, why an economy might operate out-of-equilibrium. Speaking somewhat loosely, if not all prices and wages adjust each period - whether due to existing contracts, menu costs, informational costs, etc. - then, for example, it may be in a recession that wages are "stuck" too high relative to their "natural" level resulting in unemployment "stuck" too high until such time passes that wages adjust, at which point the economy returns to long-run equilibrium.

This paper augments a typical general equilibrium model with monopolistically competitive intermediate goods firms by considering a multisector extension in which prices are fully flexible but firms face uncertainty about aggregate variables and real marginal costs. Following Makowiak and Wiederholt (2009), this uncertainty is modeled with rationally inattentive firms and results in a delay in the response of prices to monetary policy shocks. The baseline multisector

[1] This question can equivalently be put in terms of aggregate demand (why can governments manipulate aggregate demand in order to produce short-run economic effects?), aggregate supply (why is there an upward sloping short-run aggregate supply curve?), or the Phillips curve (why is there a short-run trade-off between inflation and output?).

model and an extension with relative demand shocks and intermediate production inputs provide additional targets for the attention of firms. Compared to Makowiak and Wiederholt (2009), we are able to relax the need for extreme calibration of volatilities in order to achieve a reasonable delay.

Contemporary sticky price general equilibrium models often introduce such stickiness by assuming a la Calvo (1983) that monopolistically competitive firms may only adjust their price each period with some constant exogenous probability related to the length of time since the adjustment (more generally, in these so-called "time-dependent" models the probability of adjustment may be related to the length of time since the previous adjustment). Other mechanisms include "state-dependent" models in which the probability that a firm may adjust their price depends also on the state of the economy (see for example Dotsey et al., 1999); the introduction of fixed costs associated with price-setting (see Golosov and Lucas Jr., 2007 for a recent example of this "menu-cost" approach); or the assumption that prices are fixed due to contracts (see for example Chari et al., 2000).

There exists another branch of the literature, however, that addresses the issue of short-run monetary non-neutralities by focusing instead on informational limitations of agents, so that the impediment to full adjustment is not a restriction on the flexibility to *perform* such changes, but rather a restriction on the information that would *prompt* change.[2] This idea was famously described by Phelps (1968) and Friedman (1968) and soon after was formalized by Lucas (1972) in the eponymous Lucas Islands model, in which agents face a signal-extraction problem to distinguish aggregate from idiosyncratic conditions. More

[2] See Mankiw and Reis (2010) for a summary of the recent literature.

recent incarnations of this idea can be found in Mankiw and Reis (2002), in which agents have a Calvo-like probability of receiving new information each period (the "sticky information" approach); Woodford (2001), in which agents face a dynamic signal-extraction problem; Angeletos and La'O (2010), in which the focus is on the *heterogeneity* of information imperfections across agents takes pride of place; and Makowiak and Wiederholt (2009), in which, following Sims (2003) in a a so-called "rational inattention" model, agents must split a limited amount of attention between observing aggregate and idiosyncratic conditions. It is within this literature that the current paper falls, following the rational inattention approach.

While many of the above mechanisms rely to at least some degree on an *ad hoc* imposition of sub-optimal behavior, in the seminal work on rational inattention Sims (2003) lays out a framework for microfoundations of imperfect information. While agents are still fully optimizing, in Sims' model they face an information-processing capacity constraint. Recognizing that they cannot pay attention to everything - that their information must necessarily be imperfect - they optimally use what capacity they do have.

We pursue this approach in this paper. In the perfect-information case firms set prices as a markup over nominal marginal costs, a step that requires knowledge of the contemporaneous aggregate price level, one's own contemporaneous productivity shock, and, due to the effect on aggregate demand, the contemporaneous shocks to every other firm. In the imperfect-information case, firms' uncertainty is formalized in terms of an information-processing capacity constraint, requiring firms to optimally divide their attention between aggregate and idiosyncratic shocks. In the special case of Gaussian white noise shocks, we derive analytic results on the optimal allocation of attention. More generally, the equilibrium behavior allows heterogeneous behavior related to firm- and sector-level characteristics.

This line of research is motivated by recent empirical work - in particular Bils and Klenow (2004), Golosov and Lucas Jr. (2007), Klenow and Kryvtsov (2008), and Boivin et al. (2009) - that suggests first that the results from traditional models of inducing monetary non-neutralities are not consistent with all observed inflation dynamics, and second that because disaggregated price series display markedly different inflation dynamics than do aggregated series, a successful model must begin work at the level of individual sectors. The former suggestion motivates the use of rational inattention as the key deviation allowing monetary non-neutralities - a suggestion that was also made in Sims' original work and has been already been followed up on, in particular in Makowiak and Wiederholt (2009). The latter implies that special attention must be paid to modeling sectors themselves; pursing this is one of the contributions of this paper.

Finally, this paper is especially related to three recent theoretical models. Woodford (2001), key in the recent revival of imperfect information models, describes a one-sector model in which firms face a dynamic signal-extraction problem. Whereas it *motivates* information imperfections by appealing to rational inattention, this paper (as does Makowiak and Wiederholt, 2009, below) *derives* the imperfections from optimizing behavior. Angeletos and La'O (2010) use a multisector real business cycle model to emphasize that heterogeneity of information can generate business cycles. The current paper's multisector approach and its focus on sectoral heterogeneity in particular follows from this realization. Finally, this paper can be thought of most directly as an extension of Makowiak and Wiederholt (2009) who present a one-sector rational inattention model and derive conditions for optimal allocation in the cases of Gaussian white-noise and stationary shocks. In the stationary case they describe the implications for inflation dynamics but are forced to use computational techniques. A one-sector special case of the current paper's primary result reduces to the model found in section IV of their work.

The remainder of this paper is structured as follows. In section 2, the literature is reviewed in some detail. Section 3 introduces information theory and details some useful results. Section 4 presents the model, section 5 the equilibrium conditions, and section 6 the results. Section 7 concludes.

Related Literature

With the threefold goal of (1) positioning the current paper along the arc of the existing literature, (2) explaining the relationship between this model and closely related models, and (3) describing relevant features of the data that inform modeling choices, this section proceeds by briefly describing the imperfect information literature, presenting related empirical research, and introducing recent theoretical models against which this paper's model will be contrasted.

Lucas (1972, 1973)

Robert Lucas, Jr. laid out the first formal models with imperfect information leading to monetary non-neutrality using geographically separated islands as the device preventing perfect information. Agents on each island receive signals about unknown variables and must solve a (static) signal extraction problem each period to distinguish idiosyncratic from aggregate conditions, with the key result that individuals' misperceptions of movements in nominal price for movements in real price can allow monetary policy to affect real variables. In particular, if firms (individuals in the model) mistake a purely nominal increase in the price level for an increase in their real price they will increase employment. While these simple models have been superseded, the insight that individuals' optimizing behavior depends on aggregate variables that may be unknown permeates all of the subsequent imperfect information literature, and will be a central focus in the model presented in the current paper.

Despite a high level of subsequent interest in imperfect information models, the literature largely died out by the early 1990s due especially to several critiques that could not at the time be overcome. The first was the difficulty of squaring the model with data - for example, the model required all information to become available after one period, implying that effects of monetary policy would not be more persistent than that delay. However, "periods" long enough to match the observed persistence of monetary policy effects implied an implausibly long delay before agents were made aware of that policy. Second, technical difficulties arose in modeling higher order expectations (see Townsend, 1983).

Strategic effects between individuals, a feature not highlighted in Lucas' models, have since become important and are considered in more recent imperfect information models as well as in models with other mechanisms inducing stickiness; for example, the implications of pricing decisions as strategic substitutes or as strategic complements are important to New Keynesian models (see Woodford, 2003 sections 3.1.3 - 3.1.4). Strategic considerations are similarly important in Angeletos and La'O (2010), described below, and will aid interpreting the current paper's results.

Morris and Shin (2002)

Stephen Morris and Hyun Song Shin (2002) were instrumental in restarting the discussion of imperfect information models, demonstrating welfare implications of imperfect and heterogeneous information in the presence of strategic complementarities and drawing out the link between higher order expectations and strategic behavior. In their model, agents receive public and private signals about unknown variables; one important result is that increasing precision of public signals may actually reduce welfare. While they do not focus specifically on pricing decisions, they show that Lucas' model is equivalent to the one they consider. Their intuition and solution techniques carry over to a wider range of modeling approaches, including the multisector model in Angeletos and La'O (2010) and the current paper.

Sims (2003)

Chris Sims (2003) introduced rational inattention, a modeling paradigm in which rational optimizing agents could fail to take into account even freely available information, thus providing microfoundations for information imperfections.[3] His suggestion provides a response to one critique of the Lucas model: if agents do not pay attention to monetary policy, it does not matter how quickly the information is made available. The technical component of these models introduces information theory to economics, a topic that is described below in some detail (see *Information Theory*).

Notice that many of the papers described here use "signals" received by agents as the technical device encoding imperfect information. Sims shows that

[3] For more on modeling rational inattention, see Sims (1998), Sims (2005), and Sims (2010).

in a special case - a linear quadratic optimization problem along with Gaussian stochastic processes - rational inattention can lead to results that are identical to simply assuming agents receive noisy signals, but the rational inattention approach provides a framework for the optimal selection of signals by agents and shows how the noise varies systematically with underlying model parameters.

The central contribution of the current paper is the solution of a rational inattention problem by intermediate goods firms in the presence of both idiosyncratic and aggregate shocks.

Woodford (2001)

A paper prepared by Michael Woodford for a conference commemorating Edmund Phelps was similarly important in the revival of the imperfect information literature. It extended Lucas' model to one in which information does not become available after a one period delay, so that individuals face a dynamic signal extraction problem. Agents are also less informed than in Lucas' model: not only are they unaware of aggregate conditions, they are also unaware of other agents' expectations of aggregate conditions (and their expectations of expectations of ...). Woodford shows in a one-sector model that results hinge on an infinite sum of higher order expectations and uses the Kalman filter to solve the (dynamic) signal extraction problem. The key result, driven largely by sluggishness in the response of higher-order expectations, is that the real effects of monetary policy can persist for an arbitrary number of periods.

In solving the model, the assumption is made that individuals are given signals about aggregate quantities. To justify it, Woodford briefly refers to Sims (2003) but does not explore the rational inattention problem. Pursuit of microfoundations for the optimal selection of signals by agents has been an area of subsequent research; one example is Makowiak and Wiederholt (2009) who solve the rational inattention optimal price problem for a one-sector model comparable to Woodford's. Their paper is described in detail below, and the current paper is a partial extension of their results to a multisector setting.

Angeletos and La'O (2010)

While imperfect information has traditionally been an amplification mechanism for monetary policy, Angeletos and La'O (2010) consider its ability to induce business cycles in a purely real setting. They develop a multisector model in which islands provide boundaries to information dispersion and in which intermediate goods firms set quantities (rather than prices, as has been the case above). They show that it is the heterogeneity of information across the islands rather than the magnitude of imperfection that drives their results.

They find that dispersed (and heterogeneous) information can lead to fluctuations and inertia in macroeconomic variables and that the generated fluctuations match qualitative facts about business cycles that other imperfect information models cannot (although they do not pursue any quantitative investigation). Furthermore, they emphasize that the model can generate these fluctuations even when individuals are nearly perfectly informed, so long as information is dispersed.

Strategic complementarities (although not those of the New Keynesian type, as they point out) induced by "trade linkages" (the between-island elasticity of substitution) are important in understanding the interdependence of firms' decisions and are crucial to their results. In particular, it is only when firms' decisions are complementary (goods across islands are not perfect substitutes) that imperfect information has real effects.

Their paper provides both the theoretical motivation and basic model setup for the current paper, although here we return to the consideration of price-setting firms. Their emphasis on the importance of modeling the interactions of heterogeneous agents in the presence of imperfect information provides an impetus for the extensions of Makowiak and Wiederholt (2009) that this paper considers (this point discussed at greater length below).

Mackowiak and Wiederholt (2009)

Bartosz Mackowiak and Mirko Wiederholt (2009) (hereafter MW) consider optimal price-setting behavior of rationally inattentive firms in the face of idiosyncratic and aggregate shocks, finding that for certain calibrations (in which idiosyncratic shocks are an order of magnitude more volatile than aggregate shocks) the real effects of monetary policy can persist for an arbitrary number of periods, a result that hinges on agents choosing not to pay much attention to monetary policy.

They first consider a special case of the model in which stochastic processes are Gaussian white noise and in which case an analytical result may be found. Although this case does not induce persistence in the model (since all shocks are purely transitory), it draws out the intuition of rational inattention and shows how the firms' attention allocation decision depends on model parameters. One interesting point involves strategic complementarities: strategic complementarities in price setting imply strategic complementarities in information acquisition. It is this portion of their paper that the current paper extends to a multisector setting.

Second, they consider more realistic stochastic processes for shocks; this sufficiently complicates results that computational techniques must be used to solve the information allocation problem, and show that their model can generate real effects of monetary policy and explain why firms might respond quickly to idiosyncratic shocks but slowly to aggregate shocks. In order to do this, their model requires calibrating the volatility of idiosyncratic shocks to be at least an order of magnitude larger than the volatility of aggregate shocks. While the data suggests that idiosyncratic shocks are more volatile than aggregate shocks, it does not support this degree of difference (see the discussion of the related empirical work, below, for details). One of the contributions of this paper is relaxing the differential in volatility required to achieve the appropriate behavior.

A more complete discussion of the contributions of the current paper *vis a vis* MW follows a brief summary of Makowiak et al. (2009), below.

Boivin et al. (2009)

Jean Boivin et al. (2009) use a factor augmented vector autoregression (FAVAR) approach to estimate separately the effects of aggregate and idiosyncratic disturbances on price-setting behavior, finding that prices are flexible with respect to idiosyncratic shocks but sticky with respect to aggregate (in particular monetary) shocks, a feature of the data that they suggest is not consistent with many contemporary stickiness-inducing mechanisms (for example they suggest that the Calvo mechanism cannot explain the flexibility with respect to idiosyncratic disturbances). They lay out seven stylized facts, all of which provide a rich research agenda with respect to the theoretical modeling of imperfect information, and two of which may justify modeling decisions in this paper.

First, their primary result is the importance of distinguishing between idiosyncratic and aggregate components of price change; this is of central consideration in the rational inattention approach to modeling optimal pricing decisions. Note that this result is not specific to Boivin et al. (2009) but is also borne out in related empirical efforts.[4] Second, they suggest that idiosyncratic shocks driving price changes are supply shocks at the sectoral level, providing support for the approach in the current paper for integrating idiosyncratic shocks (as opposed to introducing them as, for example, demand shocks).

Their other facts have implications for extensions of the current paper and will be of particular importance when introducing persistence in shocks. Several are suggestive of interesting extensions (for example, including a further integration of the effects of market power, see below), while several others appear to present challenges to the current approach. For example, they suggest that the reaction to all types of shocks is faster in sectors with more volatile idiosyncratic shocks, whereas the results of MW specifically suggest that increased volatility in idiosyncratic shocks reduces the reaction to aggregate shocks.

Finally, the FAVAR approach allows them to estimate the relative volatilities of the aggregate and idiosyncratic components. They find that "while the mean volatility of the common component of inflation lies at 0.33 percent, the volatility of the sector-specific component is more than three times as large". While this differential is not enough to generate the appropriate behavior in the single-sector model of MW, it can be enough in the multisector model presented here. Thus our model insulates the rational inattention approach from criticism that extreme calibrations are required.

Mackowiak et al. (2009)

Mackowiak et al. (2009) (hereafter MMW) consider a simple multisector extension to their model in MW as part of an effort to compare the ability of

[4] See in particular Bils and Klenow (2004) and Klenow and Kryvtsov (2008) for supporting empirical evidence.

several mechanisms for stickiness to match empirical results. The mechanisms they consider including the Calvo model, sticky information, menu costs, and rational inattention. The stylized facts they attempt to match are in the same vein as those described by Boivin et al. (2009), and they find that the rational inattention model is best able to fit them. There are several key points that distinguish the approach of MW and MMW from that of the current paper.

In both papers, Mackowiak et al. consider the tradeoff firms face between paying attention to aggregate conditions or idiosyncratic conditions, where idiosyncratic conditions refers specifically to the firm's own productivity shock. Productivity shocks to other firms either net to zero in aggregate, as in MW, or are simply grouped with other aggregate shocks (for example monetary policy incorporated via aggregate demand shocks), as in MWW. This leads to simpler optimization problems as firms only spread their attention between two signals and still distinguishes nominal (monetary policy) from real (productivity) shocks.

In contrast, the current paper considers firms' attention problems as between all shocks separately; this is desirable for several reasons. First, it embraces the emphasis in Angeletos and La'O (2010) of the importance of considering how interactions of heterogeneous firms with dispersed information alone can generate real effects; they note that "even if one is ultimately interested in a monetary model, understanding the positive and normative properties of its underlying real backbone is an essential first step".

Second, this paper is meant to provide a baseline model for considering more complicated models of firm interactions and allocation problems, and for introducing firm-level heterogeneity in pricing decisions. For example, it would make sense that firms are more easily able to pay attention to shocks of more closely related sectors. Integrating this requires a model like that of the current paper. Another motivating example can be found in the stylized facts of Boivin et al. (2009), who find that the speed of reaction to monetary policy shocks is related to the degree of monopoly power enjoyed by firms, an extension that will require attention to the underlying heterogeneity of firms' attention allocation problems.

De Graeve and Walentin (2014)

Two stylized facts derived in both Boivin et al. (2009) and Makowiak et al. (2009) are that aggregate shocks display substantial persistence but are characterized by low volatility, and that idiosyncratic shocks have low persistence but high volatility. It is partially these facts that validate the rational inattention approach. More recently, De Graeve and Walentin (2014) suggests that if the effects of measurement error are properly accounted for, these stylized facts may be overturned. In particular, they suggest that idiosyncratic shocks may be persistent and have low volatility, potentially presenting a problem for the rational inattention approach. The multisector model presented here allows the rational inattention approach to respond.

First, the authors note that their estimation approach does not allow them to distinguish between measurement error and additional structural shocks, but they argue that the rational inattention model in MW can not obviously account for the required additional shocks. By contrast, the multisector model developed below can easily accomodate the required additional shocks. Chapter 2 of this prospectus presents evidence that supports the features they identify as measurement error are more likely structural shocks affecting certain sectors.

Second, as discussed above, the multisector model reduces the volatility differential required by the rational inattention approach. Thus even if measurement error is distorting the stylized facts as they argue, rational inattention may still be used to explain observed price dynamics.

Information Theory

Rational inattention borrows from the theory of information and communication a mathematical model of information processing, and applies it to economic agents.[5] A telegraph wire serves as a channel through which a message passes from source as input to recipient as output. That it can only transmit a finite message in any given time interval is described as a finite Shannon "capacity".[6] In our model, an agent serves as a channel through which observations about the economy are translated into economic actions; the inability of the agent to process all information is modeled in terms of a finite Shannon capacity.

The basic quantity in information theory is entropy, a measure of the uncertainty associated with a random variable. Letting X denote a random variable with probability mass function or density P, entropy is defined as

$$H(X) = -E[\log(P(X))] \qquad \text{Entropy}$$

Notice that entropy is defined over probabilities and therefore must be positive and that is zero exactly when the distribution of X is degenerate. The units in which entropy is expressed depend on the base of the logarithm; typically "bits"

[5] See Sims (2003) or Sims (2010) for an introduction to rational inattention in economics, or Cover and Thomas (2006) for a general book length treatment of information theory. This section is largely drawn from chapters 2 and 9 of that work.

[6] The seminal work in information theory is Shannon (1948).

are used, corresponding to log base 2. Two closely related quantities are joint entropy and conditional entropy which measure, respectively, the uncertainty of two random variables together and the uncertainty of a random variable conditional on the observation of another random variable. Letting S denote a second random variable, these quantities and their connection (called the "chain rule") are defined

$$\begin{aligned} H(X,S) &= -E[\log(P(X,S))] && \text{Joint entropy} \\ H(X|S) &= -E[\log(P(X|S))] && \text{Conditional entropy} \\ H(X,S) &= H(X) + H(S|X) && \text{Chain rule} \end{aligned}$$

In the case that the two random variables are independent, the conditional entropy is identical to the (unconditional) entropy and so the joint entropy is the sum of the individual entropies.

Using these two definitions, we can define a quantity that will be of central interest in rational inattention, mutual information. The mutual information between two random variables X and S is the *reduction in uncertainty* about X given the observation of S; in that way it measures the information content contained in one variable about another. It is defined as

$$\mathcal{I}(X;S) = H(X) - H(X|S) \qquad \text{Mutual information}$$

In the case that the variables are independent so that is no reduction in uncertainty, then $I(X;S) = H(X) - H(X) = 0$. Supposing that $\mathbf{X}$ and $\mathbf{S}$ are finite n-dimensional independent vectors such that X_i and S_j are independent if and only if $i \neq j$, then

$$\mathcal{I}(\mathbf{X};\mathbf{S}) = \sum_{i=1}^{n} I(X_i; S_i)$$

This result is important because in our model we will consider attention allocation problems in which X and S will be vectors and our assumptions will make them internally independent though mutually dependent. Typically we will think of X as fundamentals of interest (for example a stochastic shock) and S as signals received by agents that provides some information about those fundamentals.
In the rational inattention framework, agents optimally choose the signals, but must do so subject to an information constraint. That constraint is formalized as a maximum level of mutual information between the fundamental and the signal: $\mathcal{I}(X;S) \leq \kappa$. Since the fundamental is a vector, in addition to respecting the overall information-capacity constraint the agents must tradeoff between paying close attention to one variable or to another.

Notice that entropy and mutual information are scalar valued regardless of the dimension of the random variables, so that all of the data regarding uncertainty and information is expressed in a single number. It is this property that leads to the simplicity of the capacity constraint, which can be introduced into the model with only a single new free parameter, κ.

Despite the ease of modeling the constraint, calibrating the value of κ is difficult for two reasons. First, it is measured in bits, where 1 bit is the level of uncertainty related to a fair coin toss. This is a difficult to interpret quantity in the context of real-world economic decisions, and is complicated further by the simplification inherent to an economic model. Second, even supposing that the bit-value of actual information-processing capacity could be assessed, the model only captures one aspect of the decision problems facing a firm, so it is unclear what proportion of their attention is specifically devoted to the stochastic elements represented in the model.

Sims (2010) suggests that when a price is assigned to additional information capacity so that the amount is variable, agents choose a relatively small amount. In practice, these models are often calibrated such that agents set prices that are close to the optimum.

Model

There are a continuum of identical households $h \in H$ with associated measure μ_H, each of which consumes a continuum of differentiated goods $j \in J$ with associated measure μ_J. Consumers have nested constant elasticity of substitution (CES) preferences that induce a partition on the goods (alternatively firms) J into sectors $\{J_1, \ldots, J_I\}$, not necessarily of equal size. For convenience and so that aggregates can be identified with averages assume the total measure of households and goods is unity, so that $\mu_H(H) = \mu_J(J) \equiv 1$. To ease the remaining notation define $\mu_i \equiv \mu_J(J_i)$ as the size (measure) of sector i in the space of all firms.[7] Sectors will be typically indexed by i; l will also be used when multiple sector-level indices are required.[8]

[7] In general a subscript j will refer to a firm $j \in J$, whereas the subscript i will denote a sector $J_i \subseteq J$. Since the sectors partition the set of goods, it is implicit that there is exactly one sector i corresponding to each firm j.

[8] The convention is that when j and i appear in the same equation, i refers to the industry such that $j \in J_i$. When a sector-level variable *not* referring to the sector of firm j is present, it will be indexed by l.

In each period households consume, supply labor to each firm j[9], buy bonds, and receive profits based on their ownership in firms. For simplicity, each household owns an equal share of every firm. Firms set their prices each period to maximize the expected value to households of profits.

Households

Household utility is a discounted stream of expected utility, additively separable in time

$$\mathcal{U}\left(\{C_{hjt}, n_{hjt}\}_{j\in J, t\geq 0}\right) = E_0 \sum_{t=0}^{\infty} \beta^t \left[U\left(\{C_{hjt}\}_{j\in J}\right) - \int_J v(n_{hjt})dj \right]$$

The nested constant elasticity of supply (CES) preferences yield two Dixit-Stiglitz aggregators. The first level of aggregation describes "sector-level" goods[10]

$$C_{hit} = \left[\int_{J_i} \mu_i^{r-1} C_{hjt}^r dj \right]^{\frac{1}{r}}$$

where the associated within-sector elasticity of substitution is $\eta = \frac{1}{1-r}$. As usual, assume that goods are gross substitutes so that $\eta \in [1, \infty)$ and so $r \in [0, 1)$. The sector-level goods are then further aggregated into a "consumption good"

$$C_{ht} = \left[\sum_{i=1}^{I} \mu_i^{1-p} C_{hit}^p \right]^{\frac{1}{p}}$$

where the associated between-sector elasticity of substitution is $\rho = \frac{1}{1-p}$. Again assuming that goods are gross substitutes yields $\rho \in [1, \infty)$ and $p \in [0, 1)$. η and

[9] This is merely for convenience. An equivalent setup has each household specializing in only one type of labor, see for example Woodford (2003) section 3.1.

[10] Notice that in the integral there appear both the indices i and j. Thus the i refers to the unique sector such that $j \in J_i$.

ρ are not constrained otherwise - for example they are not necessarily the same. Using this same basic setup, Angeletos and La'O (2010) describe the within-sector elasticity as the degree of market power of intermediate goods and the between-sector elasticity as a measure of trade linkages and strategic complementarities.

From this it is clear that the sizes of the sectors μ_i are defined entirely by consumers' relative demand weights. However, as in Woodford (2003) section 3.2.5, these weights can be reinterpreted as the product of a structural sector size parameter together with a relative demand weight.

Then given these aggregates, the households' optimization problems can be written

$$\max_{\{C_{ht}\}_{t\geq 0}\{n_{hjt}\}_{j\in J,t\geq 0}} = E_0 \sum_{t=0}^{\infty} \beta^t \left[u(C_{ht}) - \int_J v(n_{hjt}) dj \right]$$

where u is the instantaneous utility of consumption defined in terms of the consumption good and $v(n_{hjt})$ is the instantaneous disutility of labor. These are assumed to have the usual constant relative risk aversion forms

$$u(C) = \frac{C^{1-\sigma}}{1-\sigma} \quad \text{and} \quad v(n) = \frac{n^{1+\varepsilon}}{1+\varepsilon}.$$

Each period, the households' choices must satisfy the budget constraint

$$P_t C_{ht} + B_{ht+1} \leq \int_J \theta_{hj} \pi_{jt} dj + \int_J W_{jt} n_{hjt} dj + R_t B_{ht}$$

where C_{ht} is the consumption good, P_t is a price index corresponding to the cost-minimizing way to purchase one unit of the consumption good, B_{t+1} is a risk-free bond purchased in period t that yields income in period $t+1$ subject to the gross nominal risk free rate of return R_{t+1}, θ_{hj} is the share of firm j owned by the household, π_{jt} denotes profits from firm j, W_{jt} is the wage paid by firm j, and n_{hjt} is the labor provided by household h to firm j.

Firms

All intermediate goods firms produce differentiated output using a constant returns to scale technology with labor (denoted n_{jt}) as the sole input and a sector-specific productivity shock[11]

$$Y_{jt} = \varphi_{it} n_{jt}$$

For the moment we will remain agnostic about the variables present in the firms' information sets at time t. The nature of the shocks is discussed below. Assuming competitive factor markets, a firm's period profit is

$$\pi_{jt} = P_{jt} Y_{jt} - W_{jt} n_{jt}$$

Firms have a degree of market power controlled by the within-sector elasticity η, and face inverse demand curves derived from households' optimizing behavior. Thus they choose prices so as to maximize the value of their profits to the owning households whose marginal utility of wealth is $u'(C_t)$; the intertemporal problem at time t for firm j is

$$\max_{\{P_{jt+s}\}_{s=0}^{\infty}} E_{jt} \sum_{s=0}^{\infty} \left\{ u'(C_t) \left[\prod_{l=1}^{s} \frac{1}{R_{t+l-1}} \right] \left(P_{jt+s} - \frac{W_{jt+s}}{\varphi_{it+s}} \right) Y_{jt+s} \right\}.$$

[11] The model be expanded to allow a composite productivity shock, with firm-specific, sector-specific, and aggregate components, relative demand shocks, and intermediate inputs. A model with these extensions is introduced in *Extension: relative demand shocks and intermediate inputs*.

Since this model forgoes sticky prices in favor of informational frictions, firms re-optimize each period and need only solve the following static problem in each period separately

$$\max_{P_{jt}} E_{jt}\left[u'(C_t)\left(P_{jt}-\frac{W_{jt}}{\varphi_{it}}\right)Y_{jt}\right]$$

Government

Following the literature on imperfect information (Lucas, 1972, Woodford (2001), Mankiw and Reis (2002), Mankiw and Reis (2002)) we appeal to a quantity theory of money to specify a exogenous stochastic process for aggregate demand, assumed to be the result of monetary policy implemented by some policy instrument. This process will be the only aggregate shock to the economy.

$$Q_t = P_t Y_t$$

As formulated above, fiscal policy is excluded from the model to maintain focus on the firms' attention allocation problem, although in a very similar model Angeletos and La'O (2010) find that government intervention is only useful insofar is it can mitigate the distortionary effects of market power to improve efficiency, a topic not under central consideration here.

Stochastic processes

There are two exogenous stochastic processes to be specified, that for nominal aggregate demand $\{Q_t\}_{t=0}^{\infty}$ and that for idiosyncratic productivity shocks $\left\{[\varphi_{lt}]_{l=1}^{I}\right\}_{t=0}^{\infty}$. Here we assume the simple case that all processes are distributed log-normal in such a way that their logs are Gaussian white noise. All processes are assumed to be mutually independent. Formally the shocks are described

$$
\begin{aligned}
q_t &\equiv \log Q_t & \qquad q_t &\overset{iid}{\sim} N(0,\sigma_q^2) & & \\
\phi_{it} &\equiv \log \varphi_{it} & \qquad \phi_{it} &\overset{iid}{\sim} N(0,\sigma_{\phi_i}^2) & \qquad l &= 1,\ldots,I
\end{aligned}
$$

Insofar as it would be difficult to argue that independent Gaussian white noise shocks represent the true stochastic nature of the economy, this specification is only a precursor to a more complete analysis attempting to match actual macroeconomic dynamics. Unfortunately models with more realistic stochastic processes do not admit analytic results and the Gaussian white noise case provides an illuminating special case in which to consider the attention allocation trade-offs faced by firms.

One could easily accomodate non-zero mean processes by redefining the above variables as be deviations from the mean. Processes with a deterministic trend could be similarly incorporated.

For notational convenience, collect the stochastic processes into an ordered tuple

$$\Omega = \{\{q_t\}, \{\phi_{1t}\}, \cdots, \{\phi_{It}\}\}$$

indexed by ω.

Imperfect Information

The expectations operator in the above formulation of an intermediate good firm's problem suggests that at least some contemporaneous economic variables - in this case aggregate consumption, firm-specific nominal wages, the sector-specific aggregate shock, and aggregate output - are unknown to the firm at the time they must set the price of their differentiated good. This raises two questions: (1) why would a firm be unaware of these (or any) contemporaneous conditions, and (2) which contemporaneous variables are unknown to the firm. Both of these questions have been of considerable interest in the literature on imperfect information, as described above.

Here we take the position that insofar as agents must process any information they wish to use and have a limited ability to assimilate even widely available information, all variables are *a priori* unknown. Agents only observe variables at all to the extent that they specifically allocate attention to do so. This is formalized in terms of the rational inattention framework of Sims (2003) with the assumption that agents have a finite information processing capacity κ. The imperfections are thus inherent to the agents and not the information itself; in fact, this approach requires that all the relevant information exists and is freely available.[12]

Signals

The device through which agents will receive (incompletely processed) information takes the form of signals $s_{jt}^{(\omega)}$ where j is the firm receiving the signal and ω indicates one of the stochastic processes described above. While in principle the space of possible processes among which agents may select signals is unrestricted with respect to distribution, in practice the structure of this problem - in particular the Gaussian white noise exogenous processes and a log-quadratic approximation to the profit function taken below - implies that optimal signals will

[12] See Sims (2005) for a discussion of this and related topics regarding the assumptions implicit to rational attention models.

be Gaussian.[13] For this reason, we follow MW section IV in hereafter restricting the space of possible signals to those that are of the form "true state plus white noise"[14]

$$
\begin{aligned}
s_{jt}^{(q)} &= \tilde{q}_t + \psi_{jt}^{(q)} && s_{jt}^{(q)} \sim N(\tilde{q}_t, \sigma_q^2 + \sigma_{\psi_j^{(q)}}^2) && \\
s_{jt}^{(l)} &= \tilde{\phi}_{lt} + \psi_{jt}^{(l)} && s_{jt}^{(l)} \sim N(\tilde{\phi}_{lt}, \sigma_{\phi_l}^2 + \sigma_{\psi_j^{(l)}}^2) && l = 1, \ldots, I
\end{aligned}
$$

The signals are written in terms of the deviation-from-mean forms to maintain the possibility of non-zero-mean processes even though here, for example, $\tilde{q}_t = q_t - Eq_t = q_t$. Although a formal connection will be derived below, a firm's attention problem can be informally described as the optimal reduction (or more properly selection) of the noise in signals subject to a constraint on the maximum amount of noise reduction across all signals.[15]

Equilibrium

An equilibrium is a collection of processes for consumption, labor, wages, prices, and signals

$$\{C_{hjt}, n_{hjt}, W_{hjt}, P_{jt}, s_{jt}^{(\omega)}\}_{h,j,\omega,t}$$

such that markets clear, households maximize utility, and firms (1) set optimal prices given available information, and (2) direct their attention such that the

[13] See Sims (2003) or Makowiak and Wiederholt (2009) for a proof of this result.

[14] The use of the index l indicates that each firm j receives a separate signal for the shock to each industry.

[15] This model falls under the special Gaussian-linear-quadratic case. Sims (2010) section 3.2 presents this and related intuition for these types of models.

signals they receive about unknown quantities of interest are optimal. Notice that due to constant elasticity of substitution preferences, household optimization will uniquely define the processes for the aggregate price level $\{P_t\}$ and real aggregate demand $\{Y_t\}$, see (2.5) and (2.5) respectively. The full conditions governing optimal behavior are derived below.

Market Clearing

The three markets in this model - goods, assets, and labor - yield three market clearing conditions. Since we have assumed constant elasticity of substitution preferences, we need only specify market clearing in terms of the consumption good, $C_t = Y_t$. If this holds, then in equilibrium (in particular given optimal household behavior) the demand functions for intermediate and sector-level goods, detailed below, guarantee market clearing at those levels. Since this model admits a representative household, no bonds will be purchased or sold in equilibrium so that the asset market clearing condition is $B_t = 0$ for all periods t. Finally, the labor market equilibrium requires $n_{jt} = \int_H n_{hjt} dh$.

Optimal Household Behavior

Standard results for constant elasticity of substitution preferences give demand functions for disaggregated goods in terms of aggregated quantities.[16] Since all households are identical their optimal behavior will be identical and can be analyzed as derived from a optimizing representative household. For this reason, in all of what follows we drop the household subscript.

$$C_{jt} = \frac{1}{\mu_i} \left(\frac{P_{jt}}{P_{it}} \right)^{\frac{1}{r-1}} C_{it} \qquad\qquad C_{it} = \mu_i \left(\frac{P_{it}}{P_t} \right)^{\frac{1}{p-1}} C_t$$

[16] See *Constant Elasticity of Substitution Preferences* for details.

Corresponding to these demand functions are price indices prescribing the (minimal) cost of obtaining one unit of an aggregated quantity

$$P_t = \left[\sum_{i=1}^{I} \mu_i P_{it}^{\frac{p}{p-1}} \right]^{\frac{p-1}{p}} \qquad P_{it} = \left[\int_{J_i} \frac{1}{\mu_i} P_{jt}^{\frac{r}{r-1}} dj \right]^{\frac{r-1}{r}}$$

Given these demand functions, the household's intertemporal problem can be analyzed in terms only of the consumption good. As usual, the solution is described by an Euler equation and a static first-order condition[17]

$$u'(C_t) = \beta E_t \left[R_{t+1} \frac{P_t}{P_{t+1}} u'(C_{t+1}) \right]$$

$$v'(n_{jt}) = \frac{W_{jt}}{P_t} u'(C_t)$$

Optimal Price Setting

Optimal behavior on the part of the firm will be considered in two stages. First, no matter the signals they actually receive about the state of the economy, firms must optimally set their decision variable given that information. In the case of perfect information this is the standard profit maximization problem faced by a monopolist. In the case of imperfect information a log-quadratic approximation to the profit function yields the certainty equivalence result that the optimal imperfect information price is simply the expectation of the optimal perfect information price.

[17] See *Optimal Household Behavior* for details.

Second, firms must select the signals they receive. Here they achieve this by allocating their attention such that they minimize the expected loss in profits from setting a non-optimal price subject to a constraint on the maximum attention they can spread across all variables.

Perfect Information

As a baseline, consider firms with perfect information. In this case there is no attention allocation so that the firms' entire problem reduces to the standard profit maximizing problem faced by a monopolist

$$\max_{P_{jt}} u'(C_t)\left(P_{jt} - \frac{W_{jt}}{\varphi_{it}}\right) Y_{jt}$$

This yields the standard result that monopolists set price as a markup over nominal marginal costs

$$P_{jt}^{\diamond} = \frac{1}{r}\frac{W_{jt}}{\varphi_{it}}$$

which can be rewritten in terms of model fundamentals and in the form of proportional (log) deviation from the point at which all prices are the same as[18]

$$\tilde{p}_{jt}^{\diamond} = -\gamma\tilde{\phi}_{it} + \zeta\tilde{q}_t + (1-\zeta)\tilde{p}_t$$

where $\tilde{q}_t$ represents nominal aggregate demand, $\zeta \equiv \alpha(\sigma + \varepsilon)$ relates to strategic complementarities between firms' pricing decisions, $\gamma \equiv \alpha(1+\varepsilon)$, and $\alpha \equiv (1 + \rho\varepsilon)^{-1}$.[19] Integrating across all firms and applying a log-linear approximation yields

[18] See *Perfect Information*

[19] See *Optimal Price Setting*

the perfect-information equilibrium aggregate price

$$\tilde{p}_t^{\diamond} = \tilde{q}_t - \frac{\gamma}{\zeta}\sum_{l=1}^{I}\mu_l\tilde{\phi}_{lt}$$

Strategic complementarities

The firm's pricing rule (2.5) exposes $(1-\zeta)$ as a parameter governing strategic complementarities in the model. If it is positive (so $\zeta < 1$), firms' responses to changes in the aggregate price level will be complementary, whereas if it is negative, the aggregate price will act as a strategic substitute. This parameter appears in and has been important to both models with price stickiness and models with informational frictions. Typical calibrations put the value of ζ between 0.12 and 0.4, which implies that prices are strategic complements, and the parameter governing strategic complementarities, $(1-\zeta)$, is between 0.6 and 0.88.[20] This parameter will be important not only in the firms' pricing decision given available information, but also in the firms' attention allocation problem, described below.

Imperfect Information

Defining the expected value of period profits as

$$\Pi_{jt}(P_{jt}, P_{it}, P_t, Y_t, \varphi_t) \equiv E_{jt}\left[u'(Y_t)\left(P_{jt} - \frac{W_{jt}}{\varphi_{it}}\right)Y_{jt}\right]$$

firm j faces the problem $\max_{P_{jt}} \Pi_{jt}$. Taking a log-quadratic approximation to Π_{jt} around the perfect information non-stochastic equilibrium yields the following

[20] See Mankiw and Reis (2010)

formula for firm profits

$$\begin{aligned}\tilde{\Pi}_{jt} =&\hat{\Pi}_1\tilde{p}_{jt} + \hat{\Pi}_{11}\tilde{p}_{jt}^2 + \hat{\Pi}_{12}\tilde{p}_{jt}E_{jt}\tilde{p}_{it} + \hat{\Pi}_{13}\tilde{p}_{jt}E_{jt}\tilde{p}_t + \hat{\Pi}_{14}\tilde{p}_{jt}E_{jt}\tilde{y}_t + \hat{\Pi}_{15}\tilde{p}_{jt}E_{jt}\tilde{\phi}_{it}\\ &+ \text{other terms}\end{aligned}$$

where $\hat{\Pi}_1$ is a constant times the partial derivative of profit with respect to the first argument and the $\hat{\Pi}_{1*}$ coefficients are constants times the second partial derivatives, all evaluated at the point at which all prices are the same; "other terms" collects all terms of the second-order approximation that do not depend on $\tilde{p}_{jt}$ (irrelevent for our purposes since they do not affect the firm's pricing decision).[21] The associated first-order condition yields the following imperfect-information pricing rule

$$\begin{aligned}\tilde{p}_{jt}^* &= -\gamma E_{jt}\tilde{\phi}_{it} + \zeta E_{jt}\tilde{q}_t + (1-\zeta)E_{jt}\tilde{p}_t\\ &= E_{jt}\tilde{p}_{jt}^{\diamond}\end{aligned}$$

To find the imperfect information equilibrium aggregate price we follow a guess and verify approach. Given the form of the perfect-information aggregate price, we guess that under imperfect information it is described by

$$\tilde{p}_t = a\tilde{q}_t - \frac{\gamma}{\zeta}\sum_{l=1}^{I} b_l\mu_l\tilde{\phi}_{lt}$$

This guess will be verified in the next section in conjunction with the solution to the attention allocation problem. In the meantime, substituting this guess in to the

[21] See *Optimal Price Setting* and *Log-quadratic approximation to an intermediate good firm's profit function* for details.

imperfect information pricing rule yields

$$\tilde{p}^{*}_{jt} = \left[(1-\zeta)a + \zeta\right] E_{jt}\tilde{q}_t - (1-\zeta)\frac{\gamma}{\zeta}\sum_{l=1}^{I} b_l \mu_l E_{jt}\tilde{\phi}_{lt} - \gamma E_{jt}\tilde{\phi}_{it}$$

and noticing that since firms only observe Gaussian signals the expected values can be solved using typical signal extraction results:

$$\begin{aligned}\tilde{p}^{*}_{jt} &= \left[(1-\zeta)a + \zeta\right]\left(\frac{\sigma_q^2}{\sigma_q^2 + \sigma^2_{\psi_j^{(q)}}}\right) s^q_{jt} \\ &- (1-\zeta)\frac{\gamma}{\zeta}\sum_{l=1}^{I} b_l \mu_j \left(\frac{\sigma_l^2}{\sigma_l^2 + \sigma^2_{\psi_j^{(l)}}}\right) s^{(l)}_{jt} - \gamma\left(\frac{\sigma_i^2}{\sigma_i^2 + \sigma^2_{\psi_j^{(i)}}}\right) s^{(i)}_{jt}\end{aligned}$$

The firm's attention allocation problem is to select optimal signals $s^{(\omega)}_{jt}$. Since the variance of the fundamentals is given, in practice this means that firms will optimally select the variances of the signals' noise.

Optimal Attention Allocation

Having solved for optimal firm behavior given available information, we now derive the optimal information structure by considering the attention allocation problem.[22] The firm is concerned with the difference between the price it actually sets and the price it would set under full information only to the extent that it results in a loss in profits. This expected loss in profits is

$$E_{jt}\left[\tilde{\Pi}_{jt}\left(\tilde{p}^{\diamond}_{jt}, \cdot\right) - \tilde{\Pi}_{jt}\left(\tilde{p}^{*}_{jt}, \cdot\right)\right] = \left(\frac{\hat{\Pi}_{11}}{2}\right) E_{jt}\left[\left(\tilde{p}^{\diamond}_{jt} - \tilde{p}^{*}_{jt}\right)^2\right]$$

[22] See *Information Theory* and *Rational Inattention under Gaussian White Noise* for all details related to this section.

The firm's attention problem then is to optimally select signals to minimize a quadratic loss

$$\min_{\{s_{jt}^{(\omega)}\}_{\omega\in\Omega}} E_{jt}\left[\left(\tilde{p}_{jt}^{\diamond}-\tilde{p}_{jt}^{*}\right)^2\right]$$

subject to the constraint that the total information content of the signals does not exceed some value κ. This constraint can be put formally in terms of mutual information

$$\mathcal{I}\left(\left\{\tilde{q}_t,\tilde{\phi}_{1t},\cdots,\tilde{\phi}_{It}\right\};\left\{s_{jt}^{(q)},s_{jt}^{(1)},\cdots,s_{jt}^{(I)}\right\}\right)\leq\kappa$$

Then given the independence assumptions and defining for notational convenience $\kappa_j^{(\omega)} \equiv \mathcal{I}\left(\{\omega_t\};\{s_{jt}^{(\omega)}\}\right)$, it can be reformulated as $\sum_{\omega\in\Omega}\kappa_j^{(\omega)} \leq \kappa$.[23] Since the signals are Gaussian, it can be shown that the mutual information is a function only of the ratio of the variances of the fundamental and the noise

$$\kappa_j^{(\omega)}=\frac{1}{2}\log_2\left(\frac{\sigma_\omega^2}{\sigma_{\psi_j^{(\omega)}}^2}+1\right)\qquad\qquad \omega\in\Omega$$

After some algebra, the firm's attention problem can be finally written

$$\min_{\{\kappa_j^{(\omega)}\}_{\omega\in\Omega}}\sum_{\omega\in\Omega}\left(\bar{\kappa}_j^{(\omega)}\right)^2 2^{-2\kappa_j^{(\omega)}}\quad;\qquad \bar{\kappa}_j^{(\omega)}=\begin{cases}\left[(1-\zeta)a+\zeta\right]\sigma_q & \omega=q\\ (1-\zeta)\gamma\zeta^{-1}b_l\mu_l\sigma_{\phi_l} & \omega=l\neq i\\ \left[(1-\zeta)\gamma\zeta^{-1}b_i\mu_i+\gamma\right]\sigma_{\phi_i} & \omega=i\end{cases}$$

[23] Recall that ω indexes the set of all stochastic processes Ω - see *Signals* for the definition.

subject to $\sum_{\omega\in\Omega} \kappa_j^{(\omega)} \leq \kappa$ and $\kappa_j^{(\omega)} \geq 0$ for $\omega \in \Omega$. The terms $\bar{\kappa}_j^{(\omega)}$ can be thought of as *importance-weighted volatilities*; their origin is in the firm's optimal imperfect-information pricing rule (2.5).

This can be solved using standard techniques to yield the following interior solution for the optimal allocation of attention to each fundamental

$$\kappa_j^{(\omega)*} = \log_2 2^{\bar{\kappa}} + \log_2 \bar{\kappa}_j^{(\omega)} - \log_2 \bar{\kappa}_j \qquad\qquad \omega \in \Omega$$

where $\bar{\kappa} = \frac{\kappa}{|\Omega|}$ (recall that $|\Omega|$ is the number of stochastic processes) and $\bar{\kappa}_j = \left[\prod_{\omega'\in\Omega} \bar{\kappa}_j^{(\omega')}\right]^{\frac{1}{|\Omega|}}$. This equation is intuitive: the first term gives an equal allocation of attention across all stochastic processes, and the second term adds (subtracts) attention if the importance-weighted volatility of the stochastic process in question is above (below) the (harmonic) mean of importance-weighted volatility across all stochastic processes.

The above is an interior solution; we abuse notation to set

$$\kappa_j^{(\omega)*} = \begin{cases} \kappa & \kappa_j^{(\omega)*} > \kappa \\ \kappa_j^{(\omega)*} & \kappa_j^{(\omega)*} \in [0, \kappa] \\ 0 & \kappa_j^{(\omega)*} < 0 \end{cases}$$

To fully incorporate corner solutions with more than two options we need to specify a solution process in the case that $\kappa_j^{(\omega)} < 0$, which corresponds to a "negative" attention allocation to some stochastic process. This is not allowed. When this occurs, the problem can be solved by replacing the negative value with zero and then rescaling the remaining allocations such that they sum to κ.

Using these optimal attention allocations in the firms' imperfect-information pricing rules and integrating across all firms yields the imperfect-information equilibrium aggregate price

$$\tilde{p}_t^* = \underbrace{\left[\sum_{i=1}^{I} \mu_l[(1-\zeta)a+\zeta]\left(1-2^{-2\kappa_i^{(q)^*}}\right)\right]}_{a}\tilde{q}_t - \frac{\gamma}{\zeta}\sum_{l=1}^{I}\underbrace{\left[\sum_{i=1}^{I} w_{li}\left(1-2^{-2\kappa_i^{(l)^*}}\right)\right]}_{b_l}\mu_l\tilde{\phi}_{lt}$$

where $w_{li} = \left[(1-\zeta)b_l\mu_i + \zeta\mathbf{1}(l=i)\right]$ and $\mathbf{1}(l=i)$ is the indicator function that takes the value 1 if $l=i$ and is 0 otherwise. We have applied symmetry across within-industry firms to write $\kappa_j^{(\omega)^*} = \kappa_i^{(\omega)^*}\ j \in J_i,\ i=1,\ldots,I$. This verifies our guess.

Notice that through the term $\bar{\kappa}_j$, the optimal attention allocation for each fundamental depends on all of the coefficients $a, \{b_l\}_{l=1}^{I}$. For this reason, in general there are not analytic solutions to these coefficients, although computational techniques can be used to solve the fixed point problem.

Results

Interpretation

First we will consider the terms $\bar{\kappa}_j^{(\omega)}$ from the attention allocation objective function. They can be written as

$$\bar{\kappa}_j^{(\omega)} = \begin{cases} (1-\zeta)a\sigma_q + \zeta\sigma_q & \omega = q \\ (1-\zeta)\frac{\gamma}{\zeta}b_l\mu_l\sigma_{\phi_l} & \omega = l \neq i \\ (1-\zeta)\frac{\gamma}{\zeta}b_i\mu_i\sigma_{\phi_i} + \gamma\sigma_{\phi_i} & \omega = i \end{cases}$$

They are the products of structural parameters related to the importance of the process to the firm's pricing decision along with the parameter governing the volatility of the shock, and so may be considered as terms describing "importance-weighted volatility". In that light, the firm's objective is to minimize the overall importance-weighted volatility that they face.

It can be seen, moreover, that the importance of the process is derived from two distinct mechanisms related to the firm's pricing problem. First, there are direct effects: nominal aggregate demand has a direct effect on the demand curve for the firm's differentiated product and the firm's sector-specific shock has a direct effect on marginal costs. These direct effects appear as the terms that do not include the strategic complementarity $(1-\zeta)$ and do not depend on the coefficients that determine the aggregate price level $a, \{b_l\}_{l=1}^{I}$.

The second type of effect is due to strategic complementarities that arise insofar as firms are concerned with their *relative* price. If $(1-\zeta) > 0$, so that there are strategic complementarities, firms will want to raise their price in response to a general increase in prices. Larger coefficients $a, \{b_l\}_{l=1}^{I}$ amplify the effect of strategic complementarities through a larger response of the aggregate price level to firms' responses to the direct effects described above.

Finally, notice that the problem is complicated by the circularity of definitions: the firms' objective depends on the coefficients determining the aggregate price, which in turn depend on the solution to the firms' objective problem. Thus the coefficients and optimal attention levels are determined by the solution to a fixed point problem.

The optimal allocation of attention to a stochastic process can be written

$$\kappa_j^{(\omega)} = \frac{\kappa}{|\Omega|} + \log_2 \left(\frac{\bar{\kappa}_j^{(\omega)}}{\left[\prod_{\omega' \in \Omega} \bar{\kappa}_j^{(\omega')} \right]^{\frac{1}{|\Omega|}}} \right)$$

where κ denotes the total information processing capacity available to the agent, $|\Omega|$ is the total number of stochastic processes and the $\bar{\kappa}_j^{(\omega)}$ terms are as just described. The first observation is that if all importance-weighted volatility terms were equal, the second term would be zero and the optimal attention allocation would be to evenly divide capacity across all shocks. The second observation is that the denominator of the second term can be interpreted as an "average" importance-weighted volatility across all shocks. Then the optimal allocation gives more attention to those processes whose importance-weighted volatility exceeds the "average" and less attention to those that fall below the "average". Finally, notice that through the first term of the optimal allocation, as total capacity becomes infinitely large, the level of attention devoted to each shock also becomes infinitely large.

Returning to the coefficients induced by the firms' optimal allocations, notice that they can be rewritten to emphasize the strategic complementarities parameter

$$a = (1-\zeta) a \sum_{i=1}^{I} \mu_i \left(1 - 2^{-2\kappa_i^{(q)*}}\right) + \zeta \sum_{i=1}^{I} \mu_i \left(1 - 2^{-2\kappa_i^{(q)*}}\right)$$
$$b_l = (1-\zeta) b_l \sum_{i=1}^{I} \mu_i \left(1 - 2^{-2\kappa_i^{(l)*}}\right) + \zeta \left(1 - 2^{-2\kappa_l^{(l)*}}\right)$$

As noted above, this exposes the constitution of these coefficients as combinations - weighted by the strategic complementarities parameter - of the average direct effect of shocks to firms, weighted by sector size, and the effect arising through the influence of the aggregate price level on firms' *relative* prices. One interesting implication is that as strategic complementarities become strong, so that $(1-\zeta) \rightarrow 1$, the coefficient a governing the influence of monetary policy on the aggregate price level is decoupled from real aggregate demand, since that term disappears from firms' pricing rules.

Note that as total capacity tends to infinity, so that the model tends to perfectly informed agents, we have

$$\begin{aligned} a \rightarrow (1-\zeta)a + \zeta \qquad &\Longrightarrow \qquad a \rightarrow 1 \\ b_l \rightarrow (1-\zeta)b_l + \zeta \qquad &\Longrightarrow \qquad b_l \rightarrow 1 \end{aligned}$$

Thus the aggregate price under imperfect information tends to the aggregate price under perfect information as total information processing capacity becomes arbitrarily large.

A one-sector model

In the special case of a one-sector model, we have $I = 1$, $b_l \equiv 0$, and $\mu_1 = 1$. The imperfect information equilibrium aggregate price level is then

$$\tilde{p}_t^* = \underbrace{[(1-\zeta)a + \zeta]\left(1 - 2^{-2\kappa_1^{(q)^*}}\right)}_{a} \tilde{q}_t$$

and in this case, the coefficient a can be solved for explicitly

$$a = \begin{cases} \frac{(2^{2\kappa}-1)\zeta}{1+(2^{2\kappa}-1)\zeta} & \kappa_1^{(q)^*} > \kappa \\ 1 - 2^{-\kappa}\left(\frac{\gamma}{\zeta}\right)\left(\frac{\sigma_{\phi_1}}{\sigma_q}\right) & \kappa_1^{(q)^*} \in [0, \kappa] \\ 0 & \kappa_1^{(q)^*} < 0 \end{cases}$$

This is identical to the result in MW section IV, and although it appears in a slightly different form, the interpretation has the same implications as the more general discussion above.

Notice that here there is no transmission mechanism for idiosyncratic shocks to affect the aggregate price level. Although this is a one-sector model, so that the firm literally has only two signals to observe (their own productivity shock and monetary policy), the multisector extension in MWW does not depart too far from this approach in that firms still receive two signals, one regarding aggregate conditions and one regarding idiosyncratic conditions. The above discussion of the current paper's results demonstrates that there are important subtleties that arise from idiosyncratic and aggregate components of each shock, re-emphasizing the previous arguments in favor of the current paper's approach, that models each firm's attention allocation problem between each stochastic shock separately.

Calibrating volatilities

The basic point of the rational inattention approach is that firms may not immediately react to monetary policy (or other aggregate shocks) if they are not paying attention to them. Makowiak and Wiederholt (2009) formalized this concept and show that when the volatility of aggregate shocks is low relative to the volatility of a sector-specific shock, firms will optimally devote more of their attention to sector-specific conditions. If the volatility differential is great enough, the lack of attention paid to aggregate shocks will imply slow adjustment, meaning that prices will appear to be sticky in response to aggregate shocks. In their model, to achieve the appropriate degree of stickiness, they suggest that the differential in standard deviations needs to be an order of magnitude.

In particular, they calibrate idiosyncratic volatilities to match the size of average absolute price changes; this is performed under perfect information. Thus the term calibrated is the absolute expected value of $\tilde{p}^{\diamond}_{jt}$. When all disturbances are Gaussian, the absolute value is distributed half-normal with expected value $E[|\tilde{p}^{\diamond}_{jt}|] = \sigma_{p_j}\sqrt{\frac{2}{\pi}}$. Equilibrium in the one-sector and multi-sector models, along with independence assumptions, implies

$$\sigma^2_{p_j} = \sigma^2_q + \gamma^2\sigma^2_{\phi_i} \qquad \text{One-sector}$$

$$\sigma^2_{p_j} = \sigma^2_q + \gamma^2\sigma^2_{\phi_i} + \left(\frac{1-\zeta}{\zeta}\right)^2 \gamma^2 \sum_{l=1}^{I} \mu_l^2 \sigma^2_{\phi_l} \qquad \text{Multi-sector}$$

In the baseline model calibrated in MW, γ is normalized to one since changes to it have the same practical effect on the model as changes to $\sigma^2_{\phi_i}$. The calibration exercise fixes the value of σ^2_q according to the observed volatility of detrended nominal GNP, and fixes the value of $\sigma^2_{p_j}$ according to the size of average absolute price changes, as described above. Thus the volatility of the idiosyncratic shocks is fixed by their difference. The resultant calibration then has $\sigma_z = 11.8\sigma_q$.

However, empirical work suggests that while the aggregate component is less volatile than the idiosyncratic component, it is not an order of magnitude less volatile; for example, the decomposition in Boivin et al. (2009) find that for the average firm, the standard deviation of the common component (σ_q, above) is 0.33

while the standard deviation of the aggregate component (σ_z, above) is 1.09. For the median it is even closer, at 0.27 and 0.71 respectively. Even with the average values, this suggests that $\sigma_z = 3.03\sigma_q$. The effect of this change on the model is illustrated in Fig. 1.[24]

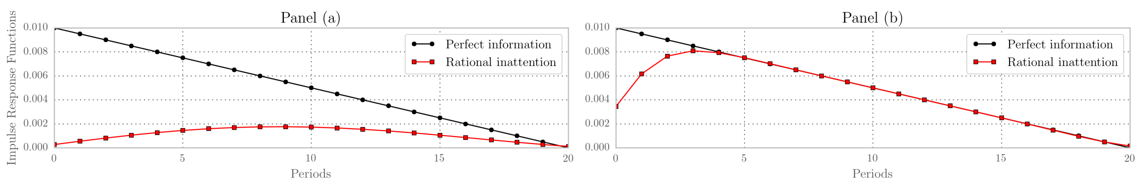

FIGURE 1. Impulse response functions in the one-sector model. Panel (a) shows the impulse responses when $\sigma_z = 11.8\sigma_q$ (as in Makowiak and Wiederholt, 2009) and Panel (b) shows the impulse responses when $\sigma_z = 3.03\sigma_q$ (as suggested by results from Boivin et al., 2009)

Furthermore, as they point out, their calibration is conservative in a number of ways. First, they exclude sales, which lowers the reported average absolute price change from 11.5% to 9.5%, and second they calibrate against the perfect information equilibrium rather than the rational inattention equilibrium. Changing either of these decisions would force their calibration to yield even more idiosyncratic volatility.

Finally, as developed in this paper, it is not merely the volatility of shocks that matters, it is also the weight in the pricing solution. This insight is important for calibrations. Here, MW normalize $\gamma = 1$, so what they calibrate as σ_z is actually $\gamma\sigma_z$. Now, if $\gamma\sigma_z = 11.8\sigma_q$, then $\sigma_z = \frac{11.8}{\gamma}\sigma_q$. Recalling that $\gamma = (1+\varepsilon)/(1+\rho\varepsilon)$ and using calibrations as in Mankiw and Reis, 2010, we calculate $\gamma \approx 1/5$, so that $\frac{1}{\gamma} \approx 5$ (alternative reasonable calibrations make γ even smaller).

[24] These figures were created using the MATLAB programs accompanying Makowiak and Wiederholt (2009) by varying the volatility calibration.

Thus their conservative calibration actually requires $\sigma_z = 59\sigma_q$. This is far removed from the results of Boivin et al. (2009).

A multi-sector model

What the analysis of the previous section points out is not that the approach of Makowiak and Wiederholt (2009) is fundamentally flawed, but rather that a more complex model is required in order to achieve realistic calibrations. Their insight that aggregate demand (i.e. monetary) shocks play a relatively small role in firms' pricing decisions merely requires that, in the language introduced above, the importance weighted volatility of aggregate demand shocks be small relative to that of other shocks.

Now the intuition for why a multisector approach is appealing is easy to see. To achieve aggregate price stickiness, all that is required is that firms pay little attention to aggregate conditions. Whereas in the one-sector case there were only two types of conditions to pay attention to (so that a decrease in attention to one meant an increase in attention to the other), here firms also pay attention to each other. This creates two channels by which an increase in idiosyncratic volatility reduces attention paid to aggregate conditions. First, firms are still concerned about their own productivity shock, which influences their marginal costs directly, and second they are concerned about the productivity shocks to all other firms because of the general equilibrium effect on aggregate demand and the aggregate price level.

In essence, the multisector model gives firms more reasons to pay attention to idiosyncratic components, which leaves less attention available for the aggregate component.

To illustrate this result, we present a simple calibration exercise for a multi-sector model, matching sectors to major groups of the Consumer Price Index (CPI). Structural parameter calibrations follow Mankiw and Reis (2010) and Nakamura and Steinsson (2010) and are presented in Table 1. The persistence and variances of aggregate demand and sectoral productivity shocks are calibrated as in MW with data on absolute price changes taken from Table VIII of Nakamura and Steinsson (2008). Weights and the absolute sizes of price changes for each sector and the resultant ratio of the variance of sectoral productivity shocks to aggregate demand shocks are presented in Table 2. The resultant volatility differentials between sectoral and aggregate fluctuations are now considerably reduced relative to the one-sector model. In fact, the weighted average of the differentials across sectors is 4.18.

TABLE 1. Benchmark structural parameters

Discount factor	$\beta = 0.99$
Coefficient of relative risk aversion	$\sigma = 1$
Inverse of Frisch elasticity of labor supply	$\varepsilon = 0$
Within-sector elasticity of substitution	$\eta = 2$
Std. dev. of innovation to aggregate demand	$\sigma_q = 0.01$
Persistence of shock processes	$\rho = 0.95$
Marginal cost of attention	1.5×10^{-4}

Using the constant marginal cost of attention approach, we numerically compute results in the case of AR(1) shocks with the above calibration. As in MW, the rational inattention parameter (here the marginal cost of attention) is chosen so that the prices are set nearly optimally.[25] Impulse response functions from two of the sectors to several types of shocks are presented in Fig. 2. Two results are apparent: (1) the multi-sector model can produce quick and strong responses

[25] As described in MW, rational inattention parameters have so far proven difficult to calibrate meaningfully. This is an important direction for future research.

TABLE 2. Sectoral calibration for multi-sector model

Name	Weight (%)	Abs. Size (%)	Volatility Differential
Processed Food	8.2	13.2	6.0
Unprocessed Food	5.9	14.2	6.5
Household Furnishings	5.0	8.7	4.0
Apparel	6.5	11.5	5.2
Transportation Goods	8.3	6.1	2.8
Recreation Goods	3.6	10.1	4.6
Other Goods	5.4	7.3	3.3
Utilities	5.3	6.3	2.9
Vehicle Fuel	5.1	6.4	2.9
Travel	5.5	21.6	9.9
Services	38.5	7.1	3.2

to idiosyncratic productivity shocks but slow and weak responses to aggregate demand shocks, and (2) it can generate heterogeneous responses across sectors to productivity shocks.

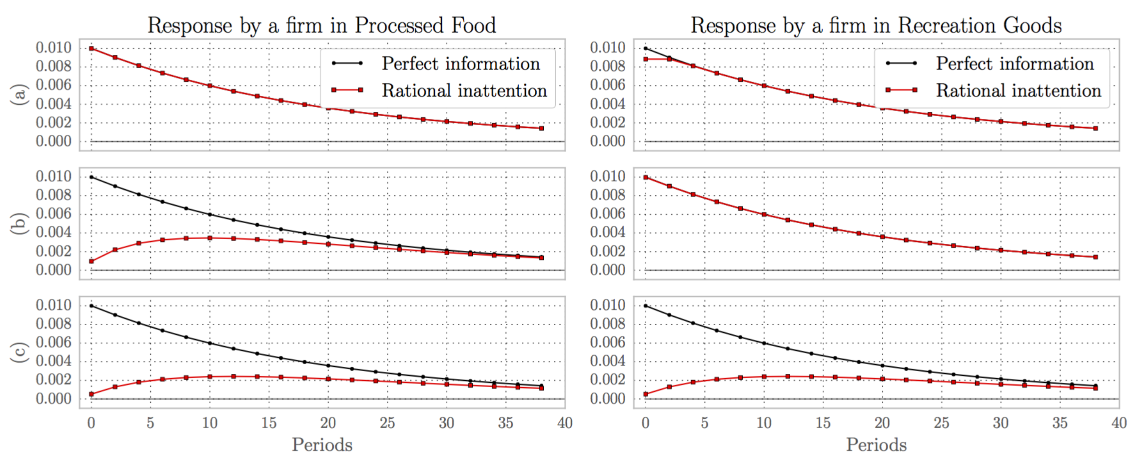

FIGURE 2. Impulse response functions in a multi-sector model. Panel (a) shows the impulse responses to a shock to the Processed Goods sector; Panel (b) shows the impulse responses to a shock to the Recreation Goods sector; and Panel (c) shows the impulse responses to a shock to aggregate demand.

In particular, in this simple illustration firms in the both the Processed Food and Recreation Goods sectors respond quickly to shocks to their own sector and they respond slowly to shocks to aggregate demand. However, while Recreation Goods firms respond quickly to shocks to Processed Foods, Processed Foods firms respond slowly to shocks to Recreation Goods.

Extension: relative demand shocks and intermediate inputs

In this section, we consider augmenting household and firm behavior to include relative demand shocks, composite productivity shocks, and intermediate inputs. These variations not only introduce desirable model characteristics but also provide additional motivation the implicit claim in the baseline model that firms pay attention to each other. Here, firms must pay attention to each other because their production process requires intermediate inputs.[26] To introduce demand shocks, we replace the demand weight μ_i with $D_{it}\mu_i$ so that the nested CES Dixit-Stiglitz aggregators can be written

$$C_{hit} = \left[\int_{J_i} (D_{it}\mu_i)^{r-1} C_{hjt}^{r} dj \right]^{\frac{1}{r}} \qquad C_{ht} = \left[\sum_{i=1}^{I} (D_{it}\mu_i)^{1-p} C_{hit}^{p} \right]^{\frac{1}{p}}$$

and where every period we require $\sum_i D_{it}\mu_i = 1$.[27] All other changes result from modifications to firms' production functions. In particular, we write:

$$Y_{jt} = \Phi_{jt} n_{jt}^{\alpha} X_{jt}^{1-\alpha}$$

where $\Phi_{ijt} = \varphi_t \varphi_{it} \varphi_{jt}$ is the composite productivity shock, and X_{jt} is a composite intermediate input constructed from the output of other firms. In particular, we suppose that, like the demand composite, it exhibits constant elasticity of

[26] See Basu (1995), Bouakez et al. (2009), and Carvalho and Lee (2011) for examples of similar models with Calvo-type pricing.

[27] Notice that this specification nests the baseline model when $D_{it} \equiv 1, i = 1, \ldots, I$.

substitution, so that

$$X_{jt} = \left[\sum_{k=1}^{I} (D_{kt}\mu_k)^{1-p} X_{ijkt}^{p} \right]^{1/p} \qquad X_{ijkt} = \left[\int_{J_k} (D_{kt}\mu_k)^{r-1} X_{ijklt}^{r} dl \right]^{1/r}$$

where X_{ijklt} is the quantity of the good produced by firm l (in sector k) used by firm j (in sector i). Similarly, X_{ijkt} is a composite of all the goods produced by firms in sector k used by firm j (in sector i). Finally, X_{jt} is a composite of the goods produced by all firms used by firm j.

Along the same lines as the baseline model, it is not too hard to show that this results in the following log-linear pricing equation[28]

$$\tilde{p}_{jt}^{\diamond} = \underline{\psi}\tilde{d}_{it} - \underline{\upsilon}\tilde{\phi}_{jt} - \underline{\gamma}(\tilde{\phi}_t + \tilde{\phi}_{it}) + \underline{\zeta}\tilde{q}_t + (1 - \underline{\zeta})\tilde{p}_t$$

where aggregate prices evolve according to

$$\tilde{p}_t = \tilde{q}_t - \frac{\underline{\gamma}}{\underline{\zeta}} \sum_{i=1}^{I} \mu_i \tilde{\phi}_{it} - \frac{\underline{\gamma}}{\underline{\zeta}} \tilde{\phi}_t$$

These equations differ from those in the baseline model through more complex parameters and the introduction of new shocks ($d_{it} \equiv \log D_{it}$ is the demand shock, and ϕ_{jt}, ϕ_t are the new productivity shock components). Qualitatively, however, it tells much the same story. Under reasonable calibrations, all of the parameters above are positive, indicating that the price set by firm j rises with relative and aggregate demand, falls with increased productivity (either firm-level, sectoral, or

[28] For example see the equation for marginal costs in Carvalho and Lee (2011).

aggregate), and, as long as there are strategic complementarities, increases with the aggregate price level.

Similarly, the rational inattention solution is qualitatively the same. More attention will be paid to those shocks that are relatively more important (the coefficient in the above pricing equation is higher) or more volatile. The log-linear rational inattention price-setting problem reduces to a sum of weighted shocks, which can be solved in the white noise case as described above, and again the unknown coefficients can be found as the solution to a fixed-point problem. Finally, it preserves the calibration result introduced above, that through the effect of multiple targets for a firm's attention prices can adjust slowly with respect to monetary policy shocks while responding quickly to idiosyncratic shocks, without requiring unrealistic volatility differentials.

Conclusion

In this paper we extend the rational inattention model of price-setting to account for multiple sectors in which firms care about their own idiosyncratic shocks, idiosyncratic shocks to other firms, and aggregate shocks. In addition to the baseline model, we consider an extension including relative demand shocks and intermediate inputs. We derive optimal attention allocations and the implied optimal price-setting behavior, allowing us to consider the effect of various parameterizations on the responsiveness of prices. The functional forms derived herein inform new directions for continued empirical research into price-setting behavior.

Our results provide several novel contributions. First, we allow firms to exhibit heterogenous behavior that depends not only on their own idiosyncratic shocks but also on firm characteristics and their relationship to other firms. Second, we show that not only can the model generate slow responses to aggregate shocks along with quick responses to idiosyncratic shocks, it can do so with less extreme parameter calibrations than in related work. Finally, we emphasize the role of importance-weighted volatility in generating optimal attention allocations, rather than volatility only.

Finally, the basic model considered here provides a baseline for future research. It would be interesting, for example, to further extend the multi-sector model to account for additional firm characteristics in order to derive testable cross-sectional implications for price-stickiness; another interesting direction is to introduce network effects as in Acemoglu et al. (2012).

CHAPTER III

ESTIMATING HETEROGENEOUS INFLATION DYNAMICS VIA UNOBSERVED COMPONENTS

Introduction

While the dynamics of aggregate inflation have been vexing economists for decades, recently it has become clear that understanding the behavior of disaggregated inflation series is key to validating theoretical models of price setting. There is wide agreement that aggregate inflation displays substantial persistence, a result that suggests that prices may be relatively "sticky". However, recent studies using micro-level data suggest that prices may change more frequently than previously thought and dynamic factor models have corroborated these results in finding that disaggregated series respond quickly and strongly to idiosyncratic shocks, but slowly and weakly to aggregate shocks. If the former shocks induce frequent price changes that cancel out in the aggregate, while the latter induce infrequent price changes that are similar across all sectors, the apparent stickiness in the aggregate price level may belie considerable flexibility in individual prices.[1] These questions are important because their answers provide guidance about the appropriate specification of price-setting behavior in theoretical models.

Unfortunately, this forward progress is complicated by the well-known issues related to collecting price data, including the presence of sales, substitutions, product or quality changes, and measurement error introduced by the sampling process. De Graeve and Walentin (2014) invoke these concerns in suggesting that

[1] See Bils and Klenow (2004) and Nakamura and Steinsson (2008) for studies using micro-level data. The seminal articles applying the dynamic factor approach are Boivin et al. (2009) and Makowiak et al. (2009)

the true variance and persistence of disaggregated inflation series are being masked by measurement errors. Refining the dynamic factor model to allow idiosyncratic processes with multiple dynamic components, they find that the data almost unanimously prefers a specification consistent with the above types of measurement error and that, properly identified, idiosyncratic shocks in fact give rise to price responses very similar to those generated by aggregate shocks.

This paper furthers the investigation into the appropriate model for the dynamic responses of prices to idiosyncratic shocks. Considering a wider range of candidate multi-component models, we find heterogeneity in the preferred specifications, most of which do not have a clear interpretation in terms of measurment error in price collection. We reassess idiosyncratic component of inflation and find that, as in the seminal paper of Boivin et al. (2009) but contrary to De Graeve and Walentin (2014), persistence is low. This then reconfirms the original result that prices appear to be flexible with respect to idiosyncratic shocks.

The paper proceeds as follows. Section 2 describes the simple and multi-component dynamic factor approaches, presents candidate models for the idiosyncratic process, and inteprets them in terms of prices and inflation. Section 3 presents analysis parallel to that considered in the previous literature, but with a wider range of candidate models, using US PCE data and discusses implications for inflation dynamics and price flexibility. Section 4 concludes.

Reduced form model

The baseline specification for a disaggregated inflation series i is the following dynamic factor model:

$$\pi_{it} = \lambda_i' C_t + e_{it}$$

$$C_t = \Phi C_{t-1} + \eta_t$$

where π_{it} is the log change in price series i, C_t is a $k \times 1$ vector of factors with associated factor loading vector λ_i, and e_{it} is a residual. The factors are extracted from a large number of time series. Together, $\lambda_i' C_t$ is termed the common component for series i, and e_{it} the idiosyncratic component. All disturbances are assumed to be Gaussian, and Φ is typically assumed to be the companion matrix for a lag polynomial, so that the factors evolve as a vector autoregression. The model is dynamic due to the evolution of the factors and because the idiosyncratic component is not constrained to be white noise. Each series can exhibit distinct dynamics through the idiosyncratic component, and also through the series-specific common component, even though all series share the same underlying driving factors, because factor loadings are series-specific. This basic model is ubiquitous in the related literature, although different approaches are used to estimate the factors and in modeling the idiosyncratic component.

Boivin et al. (2009) (BGM), De Graeve and Walentin, 2014 (DGW), and Kaufmann and Lein (2013) (KL) extract a number of factors from a large number of macroeconomic indicators including inflation series, whereas Makowiak et al. (2009) (MMW) extracts a single factor from inflation series only. In section 3, to facilitate comparison with BGM and DGW, we estimate the factors using their dataset, so that the factors are extracted from macroeconomic indicators as well as inflation series. In *Factor selection*, we show that only one of the factors extracted from the dataset of indicators explains a substantial amount of the flucuations in the disaggregated price series, and that this factor largely spans the same space as the single factor extracted from only price series. For simplicity then, we perform the estimation of section 4 using a single factor extracted from price series only.

The idiosyncratic component is modeled as an autoregressive process in all of the above cases except DGW, in which case multi-component processes are considered. Following their lead, we consider a wide range of multi-component specifications; these models are described below.

As described in, for example, Bernanke et al. (2005) and Stock and Watson (2011), there are a range of methods for estimating the factors. One of the most popular methods, used here and in all of the papers above except MMW, is principal components estimation. It is relatively easy to perform and is computationally cheap. Assuming a correctly specified model, the principal components estimator consistently recovers the space spanned by the true factors, even in the case of weak cross-sectional or serial correlation in the idiosyncratic component. Moreover, due to results in Bai and Ng (2006), the factors converge quickly enough to be used as data in subsequent estimation. This fact is used in DGW, and will be used in the results below.

Estimation proceeds in two steps. First the factors are estimated using a principal components approach. Given the factors as data, the second step puts the model into state space form and estimates both the factor loadings and any parameters governing the dynamics of e_{it} jointly via maximum likelihood estimation. The likelihood is calculated as a byproduct of the Kalman filter iterations.

The dynamic factor model, above, is a very reduced form model. Not only has it not been derived from a structural model, the factors are not structurally identified. Instead, it is a convenient way to separate the dynamics of inflation into those driven by a common component (aggregate shocks) and those driven by series-specific conditions (idiosyncratic shocks).

To get a sense of the inflation series, Fig. 3 displays three graphs. The first shows the standardized inflation series (the log change in aggregate and sectoral prices, de-meaned and normalized to a standard deviation of one). Four aggregate series, corresponding to headline CPI and the CPI series for durable goods, nondurable goods, and services, are represented by darker lines. One hundred and ninety sectoral inflation series are plotted in light blue. The second graph shows the common components (with factor loadings estimated by OLS), and the third graph shows the residual, or idiosyncratic, components. The specifics of estimating the factors will be described in section 3, below.

Idiosyncratic dynamics

Here we consider three alternative classes of specifications for the idiosyncratic component. First we describe the simple model used in Boivin et al. (2009), Makowiak et al. (2009), and Kaufmann and Lein (2013), next we describe the multi-component model of De Graeve and Walentin (2014), and finally we present the extended range of multi-component models considered here.

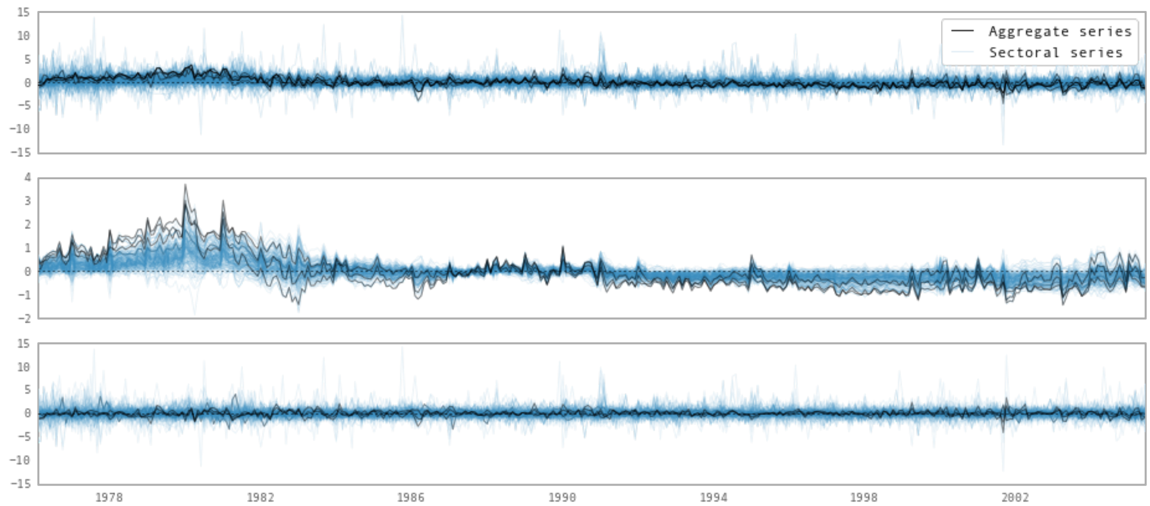

FIGURE 3. Standardized inflation series, common components, and idiosyncratic components.

The simple model

The simple model fits an autoregressive process

$$e_{it} = \rho_i(L)e_{it-1} = \varepsilon_t$$

where the number of lags is either imposed (to 13 in BGM and the simple model from DGW, to 6 in MMW, and to 1 in Reis and Watson (2010)) or is selected by information criteria (as in KL).

The refined model

The refined model of De Graeve and Walentin (2014) augments the autoregressive process with an iid white noise component, intended to control for measurement errors in inflation or item substitutions, and a moving average component, intended to control for sales or measurement errors in the price level.

The final specification is then

$$
\begin{aligned}
e_{it} &= P_{it} + I_{it} + M_{it} \\
P_{it} &= \rho_i(L)P_{it-1} + \varepsilon_{it} \\
I_{it} &= \epsilon_{it} \\
M_{it} &= \xi_{it} - \xi_{it-1}
\end{aligned}
$$

They set the number of lags to 13 for their primary analysis, but note that their results are similar if 3 lags are used or if lags are selected using standard criteria.

Anticipating the class of unobserved components models described next, it is interesting to note that this refined model can be generated as the first difference of the following local level with stationary drift model (see, for example, Clements and Hendry (2011)):

$$
\begin{aligned}
e_{it} &= \Delta y_{it} \\
y_{it} &= \mu_{it} + \xi_{it} \\
\mu_{it} &= \mu_{it-1} + P_{it} + \epsilon_{it} \\
P_{it} &= \rho_i(L)P_{it-1} + \varepsilon_{it}
\end{aligned}
$$

Generically, unobserved components models have an equivalent formulation as some ARIMA(p,d,q) specification. The ARIMA form of this model is more complicated than that of the local level model due to the drift term, but remains integrated of order one and inherits the moving average unit root.

Notice that this specification is not unreasonable for a technology process like z_{it}. As noted in DGW, their preferred interpretation, in which the multiple components account for various types of measurement errors, is observationally equivalent to an interpretation in which there simply exists a sector-specific productivity process described as a trend with stationary drift.

Unobserved components

Finally, this paper considers a wide range of candidate models, all broadly falling into the class of unobserved component models. Unobserved components models, also known as "structural time series models", are designed to explain the dynamics of a time series in terms of components with a natural interpretation, such as trends and cycles. The basic structural time series of Harvey (1990) is

$$y_t = \underbrace{\mu_t}_{\text{trend}} + \underbrace{\gamma_t}_{\text{seasonal}} + \underbrace{c_t}_{\text{cycle}} + \underbrace{\varepsilon_t}_{\text{irregular}}$$

These models are popular in deriving stylized facts of time series, since they are often more interpretable than ARIMA-type models.[2]

Notice that just as the refined model, above, nests the simple model, the class of unobserved components models nests the refined model. We thus embrace the primary conclusion of De Graeve and Walentin (2014) that the idiosyncratic process is potentially best explained in terms of multiple components. The problem of interpretation, that the specific multi-component specification in the refined model matches certain types of measurement errors as well as certain types of

[2] To clarify the somewhat confusing language, note that *trend* inflation as described by Cogley et al. (2010) and Stock and Watson (2007) is actually consistent with inflation described a *local level* model and not a local linear trend model.

structural shocks, is mitigated here since we allow the data to select between a variety of components.

The trend component allows the model to exhibit a dynamic intercept (or level), and in fact the most basic trend is just a static intercept term. The most general specification we consider allows the level to change over time and introduces a slope component allowing its rate of change to change over time as well. If both the level and the slope both evolve according to a random walk, the trend can be written:

$$\mu_{t+1} = \mu_t + \beta_t + \eta_{t+1} \qquad \eta_{t+1} \sim N(0, \sigma_\eta^2)$$
$$\beta_{t+1} = \beta_t + \zeta_{t+1} \qquad \zeta_{t+1} \sim N(0, \sigma_\zeta^2)$$

Notice that under this specification, the original time series would be integrated of order 2. If this specification was applied to prices, it would imply trending inflation; since most idiosyncratic inflation series appear to be stationary, we also consider a model that replaces the random walk evolution of the slope with an $AR(p)$ process: $\beta_{t+1} = \rho(L)\beta_t + \zeta_{t+1}$ (although of course we do not *require* a slope, or even a level term).[3] Another specialization of interest occurs when $\sigma_\zeta^2 = 0$. Then the stochastic slope is replaced with a deterministic drift term. In this case, the differenced series (inflation) would display a deterministic level (an intercept) along with an additional white noise component.[4]

[3] Given the lack of a pricing friction, this equation is not useful for serious analysis of observed pricing behavior, and is presented only to illustrate how different shocks may enter into pricing decisions. For a treatment describing how a similar model behaves in the context of a variety of pricing frictions, see Makowiak et al. (2009), section 7.

[4] See Harvey (1990) for a comprehensive treatment and Harvey and Jaeger (1993) for several specific applications. Elements of the following discussion on specification are also taken from Durbin and Koopman (2012) and Harvey and Shephard (2005).

It is important to point out one implication of the above discussion, which is that if there is a stochastic trending component at the level of prices, then at the level of inflation there will appear an additional white noise component. Therefore, if a theoretical model can result in price-setting behavior with a trending idiosyncratic fundamental shock, the I_{it} term introduced by De Graeve and Walentin (2014) has a straightforward structural interpretation.

This model also lends an interpretation to the persistent component included in both the simple model and the refined model of DGW as the slope of the price trend. If it is stationary, it suggests that unconditionally we expect the price level to evolve according to a random walk and occasionally a shock will cause an additional, persistent, drift effect that dies out over time.

A cursory examination of raw price series reveals that many have a strong seasonal component, which means that incorporating a seasonal component is important; in this literature, the data used in analysis is typically pre-processed to remove seasonal effects. In the spirit of the unobserved components approach, it would be interesting to work instead with unadjusted data and include the seasonal term directly.

Of more concern here is the possibility that the seasonal adjustment process either does not completely eliminate the seasonal effect or, in removing it, distorts other dynamic characteristics (see Harvey and Jaeger (1993) for a discussion of these issues). While we continue to use seasonally adjusted data, after we allow the possibility of an additional seasonal term. This is discussed below when we present the correlograms of the idiosyncratic series. This issue again suggests working directly with unadjusted data, a possible direction for future work.

The cycle term is intended to capture smoothly evolving and cyclical effects at frequencies much larger than those found in seasonal effects. There is no overwhelming reason to suspect this type of behavior in price-setting, but it may be present in the fundamental shocks underlying price series behavior. In practice, the data here tends not to select models with a cycle, or else estimates the cycle's frequency at the time scale of decades (indicating that the cycle is likely picking up an intercept-like feature of the data).

The irregular component is assumed to be white noise, indicating that the underlying fundamental shock is not perfectly captured by the trend, cycle, and seasonal components, but that the unexplained part has no additional structure to exploit. Notice that if this term is present at the level of prices, then at the level of inflation we would expect to see a moving average component exactly as in DGW.

Candidate models

The class of unobserved components models is very broad, and, as described above, many models can be eliminated immediately. In order to narrow the candidate models, we consider a correlogram of the idiosyncratic inflation components $\{e_{it}\}$, shown in Fig. 4 (the autocorrelations from a given series are joined with a line so that trajectories can be made out).

Several common features of most series are immediately apparent. First, for most series, the autocorrelations die away relatively quickly (with 190 series, we would expect a small number of autocorrelations to appear outside of the 95% confidence intervals each lag due to chance). This is our first indication from the data that this component of inflation may be stationary. Nonetheless a few series do appear to die away slowly, potentially indicating non-stationarities. Second, there is a negative autocorrelation at lag 1 for many series, possibly corresponding

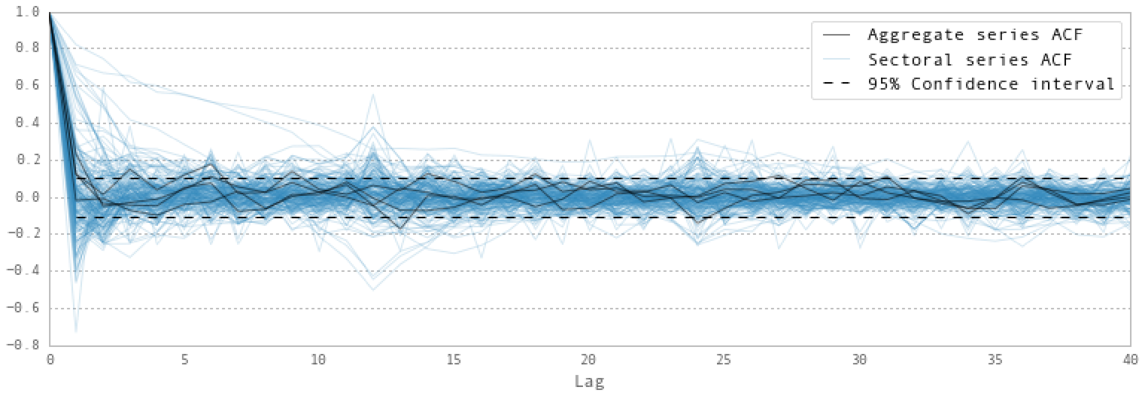

FIGURE 4. Autocorrelation functions of idiosyncratic components, 1976 - 2005

to the moving average component included by DGW and discussed above. Next, there appear to be remaining seasonal effects at lags 6, 12 and 24. Finally, the autocorrelations corresponding to aggregate series are largely within the confidence bounds, suggesting that the inflation gap (the deviation of aggregate inflation from trend) is stationary and exhibits limited persistence.

The components of the candidate models fall into the following categories: trend behavior, autoregressive persistence, and seasonal effects.

The primary candidate model for trend behavior at the price level is the local level model. Alternate specifications in the stationary case add either a deterministic or stationary drift component. All inflation series are pre-tested for the presence of a unit root, and for those series in which it cannot be rejected the drift term is modeled instead as a random walk.

Autoregressive persistence enters the model through the stationary drift term. To assess typical persistence, we consider models with up to 13 consecutive lags.[5]

[5] Stock and Watson (2007) suggest a model in which aggregate inflation is $I(1)$, but here any trend in inflation appears to be captured in the common components, rather than the idiosyncratic components.

Given that the data is seasonally adjusted, the goal is simply to prevent the effects of remaining seasonality from distorting estimates of autoregressive persistence. With this in mind, we consider models in which an unobserved components seasonal element is directly introduced into the inflation series (rather than first introduced at the price level and then differencing them to find a model for inflation, as would be appropriate if the seasonal effects themselves were of interest).

Note that the candidate models described here nest as special cases both the simple and the refined models from Boivin et al. (2009) and De Graeve and Walentin (2014). Thus, the model selection exercise in the following section allows the data to select between their models as well as the newly introduced unobserved components models, some of which are simpler and some of which have more components.

Altogether, for stationary series, we consider 8 variations in the trend / autoregressive persistence component, and 3 specifications for the seasonal component (no seasonal component, and seasonal components with periodicity 6 or 12), for a total of 24 candidate models. Note that different model specifications imply different numbers of parameters, a fact that will be important when we consider model selection.

Estimation and Results

Data

The dataset used is the same as in De Graeve and Walentin (2014) and Boivin et al. (2009). It is composed of 111 macroeconomic indicators, 4 aggregate personal consumption expenditure (PCE) inflation series, 190 sector-level inflation series and the 190 corresponding PCE quantity series, and 154 producer price index series. All series run from 1976:1 to 2005:6. As described in Appendix B to Boivin et al. (2009), all series have been transformed to induce stationarity. In the case of the price series the transformation applied was the first difference of the logarithm, so that the dataset contains inflation series.

Estimation

Estimation proceeds in two steps. In the first step, the number of underlying factors, k, is selected and the estimated factors $\hat{C}_t, t = 1, \ldots, T$ are are calculated as the first k principal components of the observed data. In all that follows, we impose $k = 5$ as in the previous literature; however, results are robust to other values, largely because regardless of the number of factors, one of the factors essentially tracks inflation behavior. This is documented in *Factor selection*. Equipped with estimates of the unobserved factors, and appealing to the results mentioned above related to the consistency of principal component estimation, we take $\hat{C}_t$ as data for subsequent analysis.

In the second step we estimate, for each series separately, the following dynamic regression model with unobserved error components:

$$\pi_{it} = \lambda_i' \hat{C}_t + e_{it}$$

where π_{it} is the i th observed PCE inflation series, λ_i' are regression coefficients, and e_{it} is one of the unobserved components specifications described in the previous section.

Since all of the candidate models (including the simple and refined models from the previous literature) fall into the class of unobserved components models, this dynamic regression model can be cast into state space form and the likelihood evaluated as a byproduct of Kalman filter iterations.[6] All parameters, including the regression coefficients, are joinly estimated via maximum likelihood estimation.

Model selection

Since all candidate models are nested as specializations of the unobserved components model, we use standard model selection criteria to select between models with the same order of integration. To select the appropriate order of integration, the augmented dickey-fuller (ADF) test for unit roots is applied as a pre-test to each series, and only the candidate models appropriate for the implied order of integration are considered. Ten series are classified as non-stationary. These series cannot be accomodated at all under either the simple or refined models, so they clearly favor the extended models presented here. These series also suggest new results in terms of persistence, since as non-stationary series, shocks are infinitely persistent.[7,8]

The specific selection criteria we consider when selecting between candidate models are the Akaike and Schwarz information criteria (AIC and SBIC, respectively) . While asymptotically equivalent, these two selection criteria can yield different results in finite samples, due to the different penalty they impose on the number of included parameters.

---

[6] See Durbin and Koopman (2012) for a book length treatment on the specification and estimation of state space models.

[7] Specifically, a series is considered to be non-stationary if the ADF test fails to reject the null hypothesis of a unit root at the 10% level. There are 10 such series. Note that if instead of pre-testing, we allow the information criteria to select between stationary and non-stationary models, 3 of these 10 series would still be classified as non-stationary using the AIC and under the BIC, 8 of the 10.

[8] A Bayesian approach to estimation would not suffer from the same difficulties regarding parameters at boundaries, like unit roots, as found in classical estimation. Thus incorporating a Bayesian approach, so that order of integration and model selection could be performed jointly, would be an interesting extension.

Generically, (see, for instance, Koehler and Murphree, 1988) either of these criteria can be written $IC = -2\log\mathcal{L} + \gamma k$ where $\mathcal{L}$ is the maximized value of the likelihood and γ is a penalty multiplier. For the AIC, $\gamma = 2$, and for the SBIC, $\gamma = \log T$. The selected model is the model that minimizes the applicable information criteria, so the effect of the larger penalty multipier is to prefer models with fewer parameters, all other things equal. Here $T = 353$, so the SBIC penalty is nearly three times as high as that for the AIC. This is important to the results presented below, since for many series there are several candidate models with similar maximized likelihood values but different numbers of parameters, and so the model selection exercise results in different specifications for roughly half of the models depending on which information criteria is used.

Since they are asymptotically equivalent, there is little to recommend one over the other, although Koehler and Murphree (1988) and Sneek (1984) suggest that the AIC has a tendency to overparameterize models. Since one goal here is to derive stylized facts about the sectoral inflation time series, we prefer parsimonious specifications that emphasize the interpretable unobserved components rather than, for example, difficult-to-interpret $AR(13)$ models. For this reason, in the discussion of stylized facts obtained by looking at the specific models selected by the data we will emphasize the results found using the SBIC.

Results related to persistence are largely unaffected regardless of the information criteria used, which suggests that it is the increased latitude to select an appropriate model, rather than the specific form of the selected unobserved components that makes our results different from those in De Graeve and Walentin (2014).

A final issue related to model selection is the time frame of the sample. Again in contrast to DGW, we find that different models are selected if we consider the post 1984 sample rather than the full sample, 1976-2005. In the context of inflation models, the mid 1980s is often associated with a substantial decrease in inflation volatility. Since stochastic volatility is not considered here, we present results from the post 1984 sample.

Table 3 presents a high level look at the results of the model selection exercise using the different information criteria. It shows the number of series for which the selected model falls into the three broad categories considered here. Since all categories of models potentially have an autoregressive persistent component, another comparison vector is the autoregressive order of selected models; these results are presented in subsequent tables.

TABLE 3. Selected models for stationary series.

	AIC	SBIC
Simple	60	65
Refined	96	36
Unobserved Components	24	79

Table 4 shows the number of models selecting each autoregressive order, across all categories, and Table 5 shows the mean and median selected lag orders broken out by category. Recall that since only stationary series are considered, the total number of series in each table is 180.

The full implications of the results will be discussed below, but a few immediate comments are merited. First, it is clear that, as described above, the additional parameter penalization in the SBIC has a dramatic effect. This is most clearly seen in the number of series for which no autoregressive component is found;

TABLE 4. Lag orders of selected models, stationary series

	0	1	2	3	4	6	12	13
AIC	4	20	28	12	22	30	46	18
SBIC	68	50	35	8	9	3	4	3

TABLE 5. Lag order statistics by model class, stationary series

(mean / median)	AIC	SBIC
Simple	4.4 / 3	2.0 / 1
Refined	8.0 / 6	3.2 / 2
Unobserved Components	5.5 / 5	0.5 / 0

this increases from 2% of the series under the AIC to 38% under the SBIC. In fact, under the SBIC, only 15% of series are well described by a lag order greater than 2.

Second, these results contrast with the model selection exercise in DGW, that finds that 88% of the series are best described by the refined model. Here the number is below 55% for the AIC, and is 20% for the SBIC. Finally, from Table 3, for both the AIC and SBIC, those series that select the refined model are also series that select longer lag lengths.

Next, we consider which components are present in the selected models under each criteria; these results are presented in Table 6 (not that rows do not sum to 100% because multiple components can be present in a single time series). The last row of Table 6 is taken from Table II in De Graeve and Walentin (2014) to show how our results differ (note that since they do not make the distinction between stationary and non-stationary series, their values take into account 10 additional series).

Since these models are at the level of inflation, “deterministic level” corresponds to the local level model with deterministic drift at the level of prices,

TABLE 6. Components of selected models, stationary series

	Deterministic level	Stationary level	Seasonal	I	M	Simple
AIC	2%	11%	7%	43%	24%	33%
SBIC	38%	6%	8%	15%	6%	36%
DGW	-	-	-	77%	66%	11%

and "stationary level" correponds to the local level model with stationary drift. I refers to the white noise component in the refined model, and M refers to the moving average component. "Simple" refers to models in which *only* an AR component is present.

It is immediately clear that regardless of the information criteria used, the presence of additional candidate models and lag orders substantially reduces the number of series for which the refined model is selected. Not only that, but, conditional on the refined model being selected, the incidence of a moving average component has declined.

Taken together, the above model selection exercise yields the following conclusions regarding the idiosyncratic components of inflation series: (1) many series are best described either an $AR(p)$ or as white noise with an intercept; and (2) across most specifications, a long autoregressive lag length is not required to fit the data well.

In terms of idiosyncratic price series, the data suggest that the local level model, or even simply an integrated autoregression of order $p \geq 0$, is a reasonable specification. This result supports the approach taken in Makowiak et al. (2009), where in the simple case the idiosyncratic component of the price follows a random walk.

Persistence

Here we use our selected models to reconsider the stylized facts of persistence in the idiosyncratic component. Figures below display histograms of the sums of autoregressive coefficients across series. Fig. 5 restricts the specification to that considered in Boivin et al. (2009), Fig. 6 restricts the specification to that considered in De Graeve and Walentin (2014), and Fig. 7 uses the model selected as above. When model selection is applicable, results are presented for both information criteria. Notice that our Fig. 5 replicates part of Figure 1 from DGW, and Fig. 6 replicates their Figure 5.

The key finding of this section is that when a wide range of candidate models is considered, the idiosyncratic inflation components display heterogeneous persistence with a median close to zero. The tables and figures just presented provide strong evidence that it is the imposition of the refined model drives up estimates of persistence and not the imposition of the simple model that drives them down.

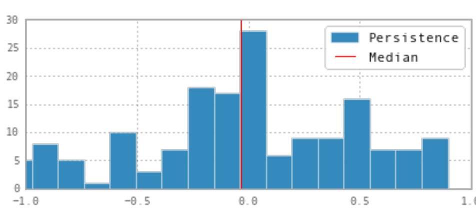

FIGURE 5. Persistence of idiosyncratic series under the simple model of Boivin et al. (2009).

In terms of validating structural models, this provides new support for the argument in BGM, MMW, and others that in order to match the facts of the data, theoretical models of price-setting must include quick and strong responses to sector-specific shocks.

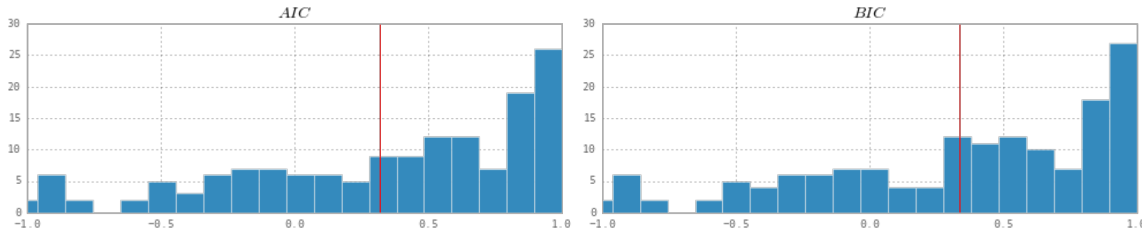

FIGURE 6. Persistence of idiosyncratic series under the refined model of De Graeve and Walentin (2014).

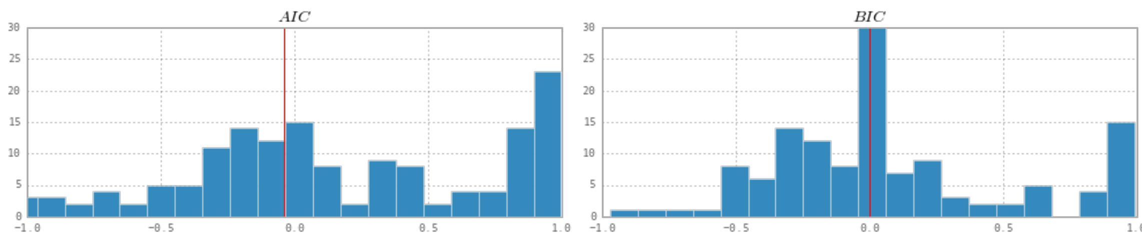

FIGURE 7. Persistence of idiosyncratic series under the class of unobserved components models considered here.

Measurement errors

Finally, let us reconsider the case for the presence of substantial measurement errors. Recall that the goal of the refined model was to use the white noise and moving average components to soak up distorting effects from substitutions and sales, respectively.

The evidence presented here suggests that this is not the role those components play. Moreover, even if measurement errors are sometimes captured by those components, there is no pervasive distorting effect. With the wider range of models considered here, the number of models in which these components entered fell substantially under both information criteria, and under the SBIC in particular they were only present in 20% of series. Thus either the data does not generally exhibit these characteristics or else their effect on model fit is small.

As described above, it is not the case that these components are wholly inconsistent with structural shocks; in fact, quite resaonable specifications for fundamental shocks could lead to similar components. The problem identified in De Graeve and Walentin (2014) is that if only the refined model is considered, the data cannot speak to the difference between structural shocks and the defined types of measurement errors. By enlarging the range of candidate models we allow the data to speak. The fact that only a moderate number of selected models contain the components supports the idea that they spring from characteristics of certain types of idiosyncratic shocks that affect only certain series, rather than from pervasive measurement issues.

As suggested by DGW, one way to assess these components is to take advantage of the characteristics of price series identified in the micro-level study of Nakamura and Steinsson (2008). They suggest that sales ought to be high in series corresponding to Apparel, Household furnishing and Food, and low in Utilities, Vehicle fuel, Services and Travel. Substitutions ought to be high in series corresponding to Apparel and Transportation goods and low in Vehicle fuel and Utilities. On the contrary, upon inspection we find that almost no apparel series have either component, that only a few food series have either component, that a number of services-related series have one or both components, and that gasoline has both components.

Conclusion

In this paper we have examined the dynamics of the idiosyncratic components of disaggregated inflation series in order to investigate the possibility of distortions caused by pervasive measurement errors and to identify stylized facts against which theoretical models of price-setting can be validated.

By embracing the multi-component approach first considered by De Graeve and Walentin (2014), but extending it to allow a wide range of candidate models, we create a setting in which the interpretation of the components can be informed by the data. Using unobserved components models, we present a unifying framework that both nests the models from the previous literature and provides a natural setting in which new models can be considered. We discuss the construction, interpretation, and implications of these models at both the level of prices and the level of inflation.

We use the model selection process and the features of the selected models to argue that rather than capturing measurement error, the unobserved components are in fact capturing the heterogeneous dynamics from fundamental shocks underlying pricing decisions by individual firms.

Finally, we reassess the stylized facts that led Boivin et al. (2009), Makowiak et al. (2009) to suggest that firms respond quickly and strongly to idiosyncratic shocks, and find that this stylized fact survives in the multi-component approach as long as the data are allowed to select heterogeneous dynamic processes, rather than being restricted to a specific imposed model.

CHAPTER IV

COMPARING APPROACHES FOR THE ESTIMATION OF DYNAMIC FACTOR MODELS

Introduction

Dynamic factor models have seen widespread use in the past few decades due to their ability to make use of the large number of relatively short time series available to macroeconomists, explain economic trends, and produce improved forecasts.

Accompanying their increased use in empirical studies has been increased attention to the theoretical and finite-sample properties of the available estimators, and there is now an extensive literature exploring non-parametric principal components estimators and maximum likelihood methods.[1] Similarly, a number of studies have explored finite-sample properties of these estimators in a variety of simulation and forecast comparison settings.[2]

In deriving theoretical properties of estimators, it is commonly assumed that the true data generating process is more complex than the one hypothesized by the estimator, so that the object of interest is consistency in large samples under misspecification; a certain robustness to misspecification is one of the attractive elements of dynamic factor models. As with any asymptotic results, however, it is not clear exactly how large a sample is required for them to obtain, particularly since there are many possible types of misspecication. Since previous work has

[1] For principal components, see especially Bai and Ng (2002), Stock and Watson (2002a), and Bai and Ng (2008b), Forni et al. (2005) for generalized principal components, and Doz et al. (2011) for a hybrid method. For maximum likelihood, see Bai and Li (2012) or Doz et al. (2012).

[2] Examples include Boivin and Ng (2005), Boivin and Ng (2006), and Alvarez et al. (2016) among many others.

found that differences between estimators may not be evident under the simple Monte Carlo analyses usually performed (see Boivin and Ng, 2005), this paper constructs a more general data generating process to explore a wider variety of misspecifications.

The combination of increased computing power and improved computational methods have made feasible estimation approaches that were previously unavailable. This paper considers two previously excluded approaches: maximum likelihood by quasi-Newton optimization and Bayesian posterior simulation by Gibbs sampling. Both methods require a more sophisticated set of computational tools, and so have not been extensively studied, although the latter method has been used in empirical work.

Section 2 describes the dynamic factor model and section 3 describes each of the estimators under consideration. Section 4 presents the data generating process used by the Monte Carlo analysis, describes the specific exercises studied, and presents the results. Section 5 concludes.

Model

Factor models relate a potentially high-dimensional dataset of observations to a small number of unobserved common factors; if the model is well-specified, the factors should be able to capture the co-movement of the observables. In particular, the literature has consistently found that the co-movements in macroeconomic time series can be explained by a few factors.[3] The use of factor models in dynamic settings was pioneered by Geweke (1976) and the dynamic factor model is now

[3] The seminal paper on this topic is Sargent and Sims (1977).

commonly written as

$$y_t = \lambda_1 g_t + \cdots + \lambda_p g_{t-p} + \varepsilon_t$$

In this formulation, the observables y_t are related to the contemporaneous and lagged values of an unobserved vector of factors g_t. The factors are themselves assumed to follow some dynamic process, often a vector autoregression. More usually in applied work, the so-called "static form" of the dynamic factor model is used. In the static form, the factors are stacked so that the observation equation can be rewritten with $\Lambda = \begin{pmatrix} \lambda_1 & \dots & \lambda_p \end{pmatrix}$ and $f_t = \begin{pmatrix} g'_t & \dots & g'_{t-p} \end{pmatrix}'$ as

$$y_t = \Lambda f_t + \varepsilon_t$$

where y_t and ε_t are $n \times 1$ vectors and f_t is an $r \times 1$ vector, for $t = 1, \dots, T$. Λ is referred to as the matrix of factor loadings, Λf_t as the "common" component, and ε_t as the "idiosyncratic" component. Specification of the idiosyncratic component will be deferred to the next section. Throughout, we will use the static form of the dynamic factor model and assume that the unobserved stacked factors f_t evolve as a vector autoregressive process.

$$\Phi(L) f_t = \eta_t$$

where the factor disturbance term satisfies $E[\eta_t] = 0, E[\eta_t \eta'_t] = Q$.

It is well known that the factors in the model described above are not separately identified from the factor loadings. In particular, for any invertible matrix A, we have $\Lambda f_t = \Lambda^* f^*_t$ where $\Lambda^* = \Lambda A$ and $f^*_t = A^{-1} f_t$. In the univariate case, this can be rephrased as the lack of identification of the "scale" of the factor. A variety of normalization strategies are available to enable parameter estimation, although in most cases the true factors will not be recovered - the estimated factors and loadings will still only be rotations of the true factors and loadings. Letting $F = \begin{pmatrix} f_1 & \cdots & f_T \end{pmatrix}'$, in this paper we normalize $F'F/T = I_r$ so that the estimators described below will recover the factors up to a rotation.[4]

Exact Factor Model

The hypothesized data generating process under which our estimators will be well-specified is the exact factor model, in which the idiosyncratic component has no cross-sectional correlation. Allowing for serial correlation, we can write the idiosyncratic component as

$$\begin{gathered} \phi_i(L)\varepsilon_{it} = e_{it} \\ \begin{pmatrix} e_{1t} & \cdots & e_{nt} \end{pmatrix}' \equiv e_t \sim N(0, H) \end{gathered}$$

Under the exact factor hypothesis, the idiosyncratic innovation covariance matrix H is diagonal.

Approximate Factor Model

For the exact factor model to hold, it must be that the common variation in the observable series can be entirely captured in the factors. Since this is unlikely to hold exactly in practice, the true data generating process is often specified to be an "approximate factor model".[5] In this case, the innovation disturbance

[4] See Bai and Ng (2013) and Bai and Li (2015) for alternative identification strategies.

[5] The designation is from Chamberlain and Rothschild (1983).

matrix H is allowed to have off-diagonal elements. Although most estimators (including those considered here) are only well-specified for the exact factor model, assumptions limiting the cross-correlation in an asymptotic sense justify their use in the approximate factor model.[6]

This paper considers the finite-sample properties of estimators when the true data generating process is assumed to be "approximate" but the hypothesis maintained by the estimators is "exact", so that the model is misspecified.

Estimators

The two objects of interest in estimating dynamic factor models are the unobserved factors (or a rotation) and the underlying model parameters. This paper considers four popular classes of estimators: principal components, maximum likelihood estimation, Bayesian estimation via posterior simulation, and a hybrid two-step estimator.

The latter three estimators require putting the model into linear Gaussian state space form, which we present here. This form consists of an "observation equation", linking the data series to the unobserved factors and a "transition equation", describing the dynamic process followed by the factors.

$$
\begin{aligned}
y_t &= \Lambda f_t + \varepsilon_t & \qquad \varepsilon_t &\sim N(0, H) \\
f_{t+1} &= \Phi f_t + \eta_t & \qquad \eta_t &\sim N(0, Q)
\end{aligned}
$$

Here we abstract from higher order vector autoregressive factor transitions and from autocorrelated idiosyncratic disturbances. In the first case, a p-th order vector

[6] See Bai and Ng (2008b) for a general discussion of these assumptions, Doz et al. (2011) for a discussion in terms of the the two-step estimator (introduced below), and Doz et al. (2012) or Bai and Li (2015) for a discussion in terms of maximum likelihood estimation.

autoregressive transition can be easily accommodated by stacking the lags of the factors and writing the transition equation as a first-order vector autoregression in companion form. In the second case, if the autocorrelation takes a known form, it can similarly be accommodated by expanding the state vector.[7]

Given system matrices Λ, H, Φ, Q, the Kalman filter and smoother (KFS) can be applied to retrieve optimal estimates of the unobserved factors.[8] Instrumental to maximum likelihood and Bayesian posterior simulation approaches, the prediction error decomposition can be used to compute the value of the likelihood function and the simulation smoother of Durbin and Koopman (2002) (or the forward-filter backward-smoother of Carter and Kohn (1994)) can be used to sample from the distribution $p(F \mid Y_T)$. The KFS can immediately accomodate missing data, and the system matrices may have arbitrary restrictions.

In addition to these benefits of adopting the state space form, forecasting is performed simply by iterating the transition equation. Letting $\hat{\Lambda}, \hat{\Phi}$ represent the estimated factor loading and VAR coefficient matrices and $f_{t|t} = E[f_t \mid Y_t]$ (the "filtered" estimate of the factor), it is easy to see that

$$\hat{y}_{t+j|t} \equiv E[y_{t+j} \mid Y_t] = \hat{\Lambda}\hat{\Phi}^j f_{t|t}$$

One vector of comparison between estimators will be the one-step-ahead forecast error.

[7] See, for example, Chapter 3 of Durbin and Koopman (2012) for a detailed presentation of state space formulations of many common time series models.

[8] In the classical framework, "optimal" here refers to minimum mean squared error estimation. Chapter 4 of Durbin and Koopman (2012) describes three other senses in which the estimates are optimal, including from the Bayesian perspective.

A potential issue that arises with conventional KFS implementations in the context of dynamic factor models (for example the presentation in Harvey (1990)) is the requirement to invert dense matrices of size $n \times n$. To make the KFS feasible for large-n models, the univariate approach of Koopman and Durbin (2000) or the collapsed approach of Jungbacker and Koopman (2008) can be used in practice. Software implementations of the KFS, often including the univariate approach, are widely available.

Principal components estimator (PCA)

The recent literature on dynamic factor models has thoroughly studied the properties of the principal components estimator; see Bai and Ng (2008b) or Bai and Ng (2013) for details. In brief, the principal components estimator is a non-parametric method that applies an orthogonal transformation to construct a few series that capture most of the covariation in the original dataset. In the factor model context, these constructed series form the PCA estimates of the factors, $\hat{f}_t^{PCA}$, and these estimates are consistent for large n and T.[9] Consistent estimates of the factor loadings are similarly available, although often the observation equation above is used to estimate the loadings via OLS from the estimated factors. To construct estimates of the VAR coefficients $\hat{\Phi}^{PCA}$, the transition equation is estimated using OLS; as usual for VAR models, this provides a consistent estimator.

The construction of the principal components and loadings implicitly imposes the normalization restrictions described in the previous section. These are sufficient that to guarantee that the estimated common component $\hat{F}^{PCA\prime}\hat{\Lambda}^{PCA}$ recovers

[9] Principal components analysis can also be viewed as a process that minimizes a residual sum of squares, see the citations in the text.

the true common component $F'\Lambda$, although separately the estimated factors and loadings are only rotations of the true factors and loadings.

It can be shown (see for example Bai and Li, 2012) that the PCA estimator coincides with the maximum likelihood estimator (described below) in the case that the idiosyncratic disturbances are homoskedastic ($H = \sigma I_n$).

Two of the great advantages of PCA are the lack of parametric assumptions and the ease and robustness of computation. The former suggests that one might expect reasonable results even in the case that the true data generating process substantially deviates from the exact factor model described above. The latter makes the PCA estimator attractive in applied work because computation amounts to calculation of eigenvalues and eigenvectors. Methods to compute these are widely available and scale well even to very large datasets. Due to the large number of parameters (models with a hundred or more observed series can easily have thousands of parameters), numerical, non-linear maximization of the likelihood function problem can prove very difficult, exhibit convergence issues, and is generally computationally expensive. On the other hand, the PCA method will rarely ever fail to produce estimates.

There are a number of disadvantages to PCA. First, PCA is not invariant to the scale of the observed data series and the ideal situation is one in which all variables share the same scale. Since this is not true in most situations, the data are usually normalized to have sample mean zero and sample variance one prior to estimation. Second, as mentioned above, it will only be efficient in the case of homoskedastic idiosyncratic disturbances; more generally, a correctly specified parametric estimator will be more efficient.[10] Finally, deviations from the standard

[10] A generalized principal components estimator allowing heteroskedasticity using an approach similar to feasible generalized least squares has been investigated, for example in Choi (2012).

model are somewhat inconvenient to incorporate. For example, known idiosyncratic dynamics, missing data, or more complex factor structures (for example the case when some factors only load on a subset of observed series) can be accommodated, but must be done on a case-by-case basis.

Two-step estimator

The two-step estimator of Doz et al. (2011) computes preliminary estimates of the factors using PCA and estimates of the system matrices by OLS in the first step, and then in the second step applies the KFS to retrieve updated estimates of the factors. As above, this approach is consistent for large n and T and incorporates some of the advantages of PCA (robust and easily computed parameter estimates) and some of the advantages of the state space form (for example ease of forecasting). One interpretation of the two-step estimator is as a single iteration of the expectation maximization algorithm, described below.

In our Monte Carlo simulations below, this will be considered the reference estimator against which the less complicated PCA estimator and more complicated maximum likelihood and Bayesian estimators will be compared.

Maximum likelihood estimation

The maximum likelihood estimator (MLE) has recently received more attention in the dynamic factor literature, see especially Doz et al. (2012), Bai and Li (2012), and Bai and Li (2015). Since the true data generating process is the approximate factor model but the maintained hypothesis for estimation is the exact factor model, MLE is usually interpreted as a quasi-maximum likelihood estimator in the sense of White (1982).

In order to compute the likelihood function the model is put into state space form and the prediction error decomposition is applied. To be concrete, the objective function is[11]

$$\log \mathcal{L}(\theta \mid Y_T) = -\frac{Tn}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \left(\log |F_t| + v_t' F_t^{-1} v_t \right)$$

where θ collects the parameters from the system matrices and v_t, F_t are the one-step-ahead prediction error and covariance matrix computed by the KFS. This optimization problem does not have an analytic solution, so numeric methods must be used. As shown in the references above, the resulting quasi-MLE estimator is consistent.

There are two available strategies for numerically maximizing the likelihood function of a dynamic factor model cast into state space form.[12] The first is applying a quasi-Newton algorithm (for example the widespread BroydenFletcherGoldfarbShannon, or BFGS, method) and the second is applying the expectation maximization (EM) algorithm of Dempster et al. (1977) and Watson and Engle (1983).

These two methods are complimentary. While the EM algorithm is generally robust even to very poor starting parameters (and under certain assumptions is guaranteed to increase the likelihood at each iteration), it can take a large number of iterations to converge to the optimum. While the quasi-Newton method is less robust to poor starting parameters and suffers to a greater extent from the curse of dimensionality, when starting close to the optimum it converges quadratically. For

[11] See Durbin and Koopman (2012), Chapter 7.

[12] Like many numerical maximization routines, the algorithms described here generally consider local maximization.

this reason, it is often recommended that the EM algorithm be initially applied for some fixed number of iterations and then a quasi-Newton method be applied for the remaining iterations until convergence is obtained.

Although quasi-Newton methods are more popular than the EM algorithm for many types of models, they often require numeric computation of the score vector which can be prohibitively time-consuming to compute in large-dimensional models, even with the recent improvements to the KFS, because the Kalman filter must be run once per parameter at each iteration. For this reason, they have received little attention in the high-dimensional dynamic factor literature, and the EM algorithm is usually used by itself.[13] Kose et al. (2003) partially motivate their choice of Bayesian methods due of the impracticality of using numerical derivatives in a quasi-Newton scheme.

Nonetheless, it is possible to compute the score vector analytically using a single pass of the KFS as shown in Koopman and Shephard (1992). Jungbacker and Koopman (2008) and Jungbacker et al. (2011) use the analytic score along with a quasi-Newton method to estimate dynamic factor models. A contribution of this paper is considering the effect of applying both the EM and quasi-Newton methods as described above.

To facilitate comparison of the estimators, here we briefly describe the intuition of the EM algorithm as applied to the model (4.3), in the case of no missing data.[14] First, notice that if the factors were known the observation and transition equations could separately be consistently estimated by OLS. This was the approach of the non-parametric estimators, where the PCA estimates of the

[13] An influential paper using the EM algorithm alone is Reis and Watson (2010).

[14] A more complete recent treatment for dynamic factor models can be found in Babura and Modugno (2014).

factors were used in place of the true factors. The EM algorithm essentially iterates on this process: at each iteration, estimates of the factors are first constructed and then the parameters are (re-)estimated to maximize the "expected" likelihood function, in which the unknown factors are replaced by their expectation. These two steps are known respectively as the expectation and maximization steps. In this case, given the the known factors, the expected likelihood function can be maximized analytically; as might be expected, the optimal parameters are the least squares estimates.

Practically, then, the EM algorithm consists of iterations in which the KFS is first applied to compute the "smoothed" factors $\hat{f}_t \equiv E[f_t \mid Y_T]$, and then the parameters of the observation and transition equations are estimated by OLS. The EM algorithm requires starting parameters to begin the iterations; a natural (and often used) set of starting parameters are those constructed by applying OLS to the PCA factor estimates - i.e. by applying the two-step estimator. Iterations continue until the differences in the likelihood in subsequent steps fall below a given tolerance.

Bayesian estimation

Bayesian estimation of dynamic factor models has generally received less attention in both the theoretical and applied dynamic factor literature. Two of the most influential papers that apply the Bayesian approach are Kose et al. (2003) and Bernanke et al. (2005). While the former uses it exclusively, the latter suggests that the PCA estimator outperforms the Bayesian estimator to some degree. Similarly to MLE, the Bayesian estimator here will be misspecified but will retain desirable asymptotic properties; for example, the posterior will be centered at the MLE estimates.[15]

Here we give only the briefest overview of Bayesian methods as applied to state space models; complete treatments can be found in Koop (2003) or West and Harrison (1999).

The Bayesian approach to parameter estimation begins by considering parameters as random variables and applying Bayes' theorem to derive a distribution for the parameters conditional on the observed data. In this case, the posterior is not available analytically so we simulate the posterior using Markov chain Monte Carlo (MCMC) techniques and apply a law of large numbers so that sample averages can be used to approximate population quantities.

Selection of a prior distribution for the parameters is required in Bayesian applications; here we select natural conjugate priors for two reasons. First, dynamic factor models are generally characterized by large amounts of data so that the effect of the prior on the posterior will typically be small, and second the Gibbs sampling scheme made possible by these priors yields a form that makes clear the relationship with the other estimation methods discussed above. Writing the observation equation line-by-line as $y_{it} = \lambda_i f_t + \varepsilon_{it}$ with $\varepsilon_{it} \sim N(0, \sigma_i^2)$, we select independent Normal-Gamma priors: $\lambda_i \mid \sigma_i^2 \sim N(0, I_r)$ and $\frac{1}{\sigma_i^2} \mid \lambda_i \sim \Gamma(10^{-4}, 3)$. Considering the transition equation as a seemingly unrelated regression with fixed variance (due to the identification strategy described above), we select a restricted Normal prior so that $vec(\Phi) \sim N(0, I_{r^2})_{[\rho(\Phi)<1]}$ where $\rho(\cdot)$ is the spectral radius and the restriction guarantees that the factors are stationary. These priors are similar to those selected in Bernanke et al. (2005).

[15] See, for example, Mller (2013) for a discussion of misspecification in the Bayesian context.

Because we have selected conjugate priors, the MCMC method used in this paper is a Gibbs sampling algorithm as in Carter and Kohn (1994). For other prior selections that do not permit Gibbs sampling, Metropolis-within-Gibbs schemes could alternatively be used with only minor modifications.

Again to facilitate comparison of the estimators, we briefly describe the intuition of the Gibbs sampler. Gibbs sampling allows the simulation of the full (joint) posterior through simulation of a collection of k conditional posteriors $\left\{\pi\left(\theta^{(i)} \mid Y_n, \theta^{(-i)}\right)\right\}_{i=1}^{k}$, where $\theta = \cup_{i=1}^{k}\theta^{(i)}$, $\cap_{i=1}^{k}\theta^{(i)} = \emptyset$, and $\theta^{(-i)} = \theta \setminus \theta^{(i)}$. The partitions of the parameters are referred to as blocks.

The Gibbs sampler is an iterative algorithm that proceeds as follows. Given a vector parameters drawn from the posterior, $\theta_{j-1}^{(-1)}$, new parameters are drawn block-by-block.[16] First, $\theta_j^{(1)}$ are drawn from $\pi\left(\theta^{(1)} \mid Y_n, \theta^{(-1)} = \theta_{j-1}^{(-1)}\right)$. As described in the references above, this constitutes a valid draw of the $\theta^{(1)}$ block from $\pi(\theta \mid Y_T)$. Then $\theta_j^{(2)}$ are drawn from $\pi\left(\theta^{(2)} \mid Y_n, \theta^{(1)} = \theta_j^{(1)}, \theta^{(-1,2)} = \theta_{j-1}^{(-1,2)}\right)$; this continues until all blocks have been drawn, at which point iteration j is complete and the collection $\cup_{i=1}^{k}\theta_j^{(i)}$ is a valid draw from the full (joint) posterior.

With the dynamic factor model cast into state space form, we partition the parameter vector into four blocks: $\theta = \{F, \Lambda, H, \Phi\}$. To construct the conditional posterior distributions, notice that conditional on the factors F, the observation equation reduces to a line-by-line linear regression and the transition equation is a seemingly unrelated regression (SUR). A standard result applying conjugate priors to regression models yields independent Normal-Gamma conditional posteriors in the line-by-line case and independent Normal-Wishart conditional posteriors in the

[16] The immediate problem is how to start the chain, since an initial draw from the posterior is not available. As usual with MCMC methods, the iteration can be initialized with arbitrary starting parameters but then requires a "burn-in" period in which the Markov chain converges to the posterior.

SUR case. Then given the parameters of the model, sampling from the conditional posterior of the factors is possible through the KFS simulation smoother. Thus the Gibbs sampler proceeds as follows:

0. Initialize the sampler with starting parameters.
1. Draw F_j from $\pi(F \mid Y_T, \Lambda = \Lambda_{j-1}, H = H_{j-1}, \Phi = \Phi_{j-1})$ using the simulation smoother.
2. Draw Λ_j line-by-line from $\pi(\Lambda \mid Y_T, F = F_j, H = H_{j-1}, \Phi = \Phi_{j-1})$
3. Draw H_j line-by-line from $\pi(H \mid Y_T, F = F_j, \Lambda = \Lambda_j, \Phi = \Phi_{j-1})$
4. Draw Φ_j from $\pi(\Phi \mid Y_T, F = F_j, \Lambda = \Lambda_j, H = H_j)$
5. Repeat steps 1-4 until enough draws have been taken from the converged posterior.

Intuitively, step (1) performs a function synonymous with the expectation step of the EM algorithm; it provides an "estimate" of the factors. Steps (2-4) perform a function synonymous with the maximization step; conditional on the factors, they "estimate" the parameters. Except for the influence of the prior (which is asymptotically negligible), the conditional posteriors in steps (2-4) are centered on the OLS estimates. This correspondence provides an intuitive explanation of the way in which one difference between the Bayesian and ML estimators - that in the former a full distribution is constructed whereas in the latter only a point estimate is found - is exhibited in the dynamic factor case.[17]

[17] See also footnote 4 of Otrok and Whiteman (1998) for a description of the relationship between the EM algorithm and Gibbs sampling.

Monte Carlo Exercises

In order to better understand how factor and parameter estimation are affected by complex features of real-world data, we investigate the following elements: (1) inclusion of unrelated data series; (2) outliers; and (3) more complex factor transitions. In addition, we estimate baseline models for comparison. A short description of these features is provided following the presentation of the data generating process, which includes serial and cross-sectional correlation in idiosyncratic disturbances.

Data generating process

This section describes the generic data generating process for the Monte Carlo simulations. First, we decribe how the observed series are constructed from the factors

$$
\begin{aligned}
y_{it} &= \sum_{j=1}^{r} \Lambda_{ij} f_{it} + e_{it} && i \in N_1 \cup N_2;\ j = 1, \dots r \\
y_{it} &= e_{it} && i \in N_3 \\
\Lambda_{ij} &\sim N(0,1) && i \in N_1 \cup N_2;\ j = 1, \dots, r \\
\Lambda_{ij} &= 0 && i \in N_3;\ j = 1, \dots, r
\end{aligned}
$$

The observed data series, denoted y_{it} are divided into into three groups; the first and second groups ($N_1 = \{1, \dots, n_1\}, N_2 = \{n_1 + 1, \dots, n_2\}$) contain data series that are generated by the factors, whereas the third group ($N_3 = \{n_1 + n_2 + 1, \dots, n_1 + n_2 + n_3\}$) is unrelated to the factors. Next, we specify the dynamic processes followed by the factors and idiosyncratic disturbances.

$$
\begin{aligned}
\Phi(L)f_t &= u_t & u_t &\sim t_\nu(0, Q) \\
D(L)e_t &= v_t & v_t &\sim t_\mu(0, \mathcal{T}) \\
\Phi(L) &= I_r - A_1 L - \cdots - A_p L^P \\
D_{ij}(L) &= \begin{cases} 1 - dL & i = j \\ 0 & \text{otherwise} \end{cases} & & i, j \in N_1 \cup N_2 \cup N_3
\end{aligned}
$$

The persistence of idiosyncratic series is controlled by the autoregressive parameter d. The vector autoregressive coefficient matrices $\{\Phi_i\}_{i=1}^{p}$ are constructed by generating matrices $\{B_i\}_{i=1}^{p}$ and applying the transformation described in Ansley and Kohn (1986). Each entry of each matrix B_i is distributed $N(0, \xi^2)$. The parameters t_ν and t_μ control the quantity of "outliers" generated by the model (lower values imply fatter tails and increased incidence of outliers), and the parameter ξ^2 acts to control persistence (the higher is ξ^2, the more persistent on average is the resultant VAR process). Next we specify the generation of the idiosyncratic innovation correlation structure.

$$
\begin{aligned}
\sigma_i^2 &= \begin{cases} \sum_{j=1}^{r} \Lambda_{ij}^2 & i \in N_1 \cup N_2 \\ \sigma^2 & i \in N_3 \end{cases} \\
\beta_i &\sim \mathcal{U}\left([u_\beta, 1 - u_\beta]\right) & i \in N_1 \cup N_2 \cup N_3 \\
\alpha_i &= \sigma_i^2 \frac{\beta_i}{1 - \beta_i} \frac{1}{1 - \rho(A)^2} \\
\mathcal{T}_{ij} &= \left[(1 - \mathbf{1}(i, j))\tau\right]^{|i-j|} (1 - d^2)\sqrt{\alpha_i \alpha_j}
\end{aligned}
$$

where the indicator function is

$$\mathbf{1}(i,j) = \begin{cases} 1 & (i \in N_1 \;\&\; j \notin N_1) \mid (j \in N_1 \;\&\; i \notin N_1) \\ 0 & \text{otherwise} \end{cases}$$

The parameter σ^2 acts like variance multiplier on the “unrelated” series, u_β controls heteroskedasticity, and τ controls the extent of cross-correlation. Finally we specify the generation of the factor innovation correlation structure.

$$\begin{aligned} \delta_i &\sim \mathcal{U}\left([u_\gamma, 1-u_\gamma]\right) && i = 1, \ldots, r \\ \gamma_i &= \frac{\delta_i}{1-\delta_i}\frac{1}{1-d^2} \\ Q_{ij} &= q^{|i-j|}(1-\rho(A)^2)\sqrt{\gamma_i\gamma_j} \end{aligned}$$

The parameter u_γ controls heteroskedasticity, and q controls the extent of cross-correlation.

By tuning the free parameters, this framework can incorporate the desirable features described above to a greater or lesser extent. The free parameters are collected and described in Table 7.

Exercises

We first consider a baseline specification replicating known results and then introduce complicating features individually. Finally, we consider a model with all features included simultaneously.

TABLE 7. Free parameters in Monte Carlo data generating process

Parameter	Description
$\%n_1$	Percentage of series that are "clean"
$\%n_2$	Percentage of series that are "contaminated"
$\%n_3$	Percentage of series that are "unrelated"
r	Number of unobserved factors
p	Order of factor lag polynomial
ν	Degrees of freedom of factor innovation t-distribution
μ	Degrees of freedom of idiosyncratic innovation t-distribution
d	Idiosyncratic autoregressive coefficient
ξ	Controls persistence of the factor lag polynomial; $\xi \to 0$ yields zero persistence
σ^2	Variance multiplier for "unrelated" series
u_β	Controls heteroskedasticity of idiosyncratic disturbances; $u_\beta \to 0.5$ yields homoskedasticity
τ	Controls cross-correlation of idiosyncratic disturbances; $\tau \to 0$ yields independence
u_γ	Controls heteroskedasticity of factor disturbances; $u_\gamma \to 0.5$ yields homoskedasticity
q	Controls cross-correlation of factor disturbances; $q \to 0$ yields independence

Baseline

The baseline data generating process corresponds to the Monte Carlo exercise presented in Doz et al. (2012), which is a specialization of the data generating process presented above. In particular, by setting $n_2 = n_3 = 0$, $\nu = \mu = \infty$, $Q = I_r$, $p = 1$, and $A_1 = 0.9I_r$, we recover their model. The full specification is provided in Table 8. For all other exercises, a specification table will be presented that will only include the parameters that differ from the baseline specification.

Because $d > 0$, $u_\beta \neq 0.5$, and $\tau > 0$, the baseline model contains heteroskedasticity and both serial- and cross-correlation. The factors are highly persistent and independent from each other. Although this falls into the class of approximate factor models, it is comparatively well-behaved.

TABLE 8. Baseline parameter specification

Parameter	$\%n_1$	$\%n_2$	$\%n_3$	r	p	ν	μ	d	ξ	σ^2	u_β	τ	u_γ	q
baseline: AR(1)	100%	0%	0%	1	1	∞	∞	0.5	-	-	0.1	0.5	-	-
baseline: VAR(1)	100%	0%	0%	3	1	∞	∞	0.5	-	-	0.1	0.5	-	-

Outliers

In many studies making use of dynamic factor models, a small number of outlier observations are identified and eliminated prior to estimation. In order to generate models with outliers, here we generically specify innovations as $t-$ distributed. By setting an infinite number of degrees of freedom we can recover the typical Gaussian case, but we can also investigate the effect of outliers by specifying a finite number.

We consider three specifications incorporating outliers, with full details in Table 9. The first specifies that the idiosyncratic disturbances are Cauchy distributed ($\nu = 1$), the second specifies a t distribution with $\nu = 2$, and the third specifies that both idiosyncratic and factor disturbances are $t-$ distributed with degree of freedom $\nu = \mu = 10$.

For the first two exercises, we also consider a variant of the estimation procedure in which outliers, defined to be data points more than 4 standard deviations from the mean, are replaced with the sample mean. Some variant of this procedure is commonly done in empirical studies.[18] In large samples, approximately 16% of innovations would be classified as outliers under the Cauchy specification, about 6% when the degree of freedom is 2, and 0.25% when the degree of freedom is 10. With infinite degree of freedom (the Gaussian case), only 0.006% would be so classified.

Unrelated series

Although many studies making use of dynamic factor models use standard large datasets, recent work suggests that such models are not immune to data

[18] See for example Jungbacker and Koopman (2008) and Reis and Watson (2010).

TABLE 9. Changed parameter specifications for outlier exercises

	ν	μ	Remove outliers
outliers: cauchy	1	∞	No
outliers: cauchy + removal	1	∞	Yes
outliers: dof=2	2	∞	No
outliers: dof=2 + removal	2	∞	Yes
outliers: dof=10	10	10	No

selection issues.[19] As pointed out by Boivin and Ng (2006), Bai and Ng (2008a), and Alvarez et al. (2016), incorporating large numbers of series that are unrelated to the factors of interest may reduce estimation performance.

The above data generating process allows for three types of series. "Clean" series (those in N_1) are the types typically studied in dynamic factor Monte Carlo exercises; they are related to the factors and exhibit a small amount of cross-correlation with each other. Following Boivin and Ng (2006), we include two other types of series. "Unrelated" series (those in N_3) are unrelated to the factors, although they may follow an autoregressive process and may be correlated with each other. Finally, "contaminated" series (those in N_2) are generated by the factors but are correlated with the "unrelated" series.

The goal of these exercises is to gauge the extent to which increased numbers of unrelated series will make it more difficult to recover the underlying factors and the extent to which more volatile unrelated series degrade forecasting perfomance. We consider three specifications incorporating unrelated data series, with parameterizations provided in Table 10.

[19] Due to frequent use, the Federal Reserve Economic Database has even recently made available ongoing updates to the dataset popularized by Stock and Watson (2002b).

TABLE 10. Changed parameter specifications for unrelated exercises

	$\%n_1$	$\%n_2$	$\%n_3$	σ^2
unrelated(20%, 10)	40%	40%	20%	10
unrelated(60%, 1)	20%	20%	60%	1
unrelated(60%, 10)	20%	20%	60%	10

Factor transition dynamics

While it is common in empirical work to specify factors as having a vector autoregressive transition (see for example Bernanke et al. (2005) and Boivin and Ng (2005) among many others), Monte Carlo studies typically have factors evolving according to univariate autoregressions, often of order one. Since an important question is whether explicitly specifying the transition dynamics through the state space form improves efficiency in finite samples, it is important that we allow more complex transition dynamics. In particular, we allow for varying degrees of persistence, varying lag orders, and vector autoregressive processes.

Specifically, we consider four parameterizations. The first two only modify the autoregressive coefficient to investigate the effect of persistence. The last two exercises instead draw the persistence randomly as described above. Of those, the first incorporates a higher lag order and the second incorporates both an increased lag order and vector autoregressive dynamics. Specifications are in Table 11.

TABLE 11. Changed parameter specifications for factor transition exercises

	r	p	ξ	u_γ	Alternative transition
AR(1) low persistence	1	1	-	-	$A_1 = 0.2$
AR(1) high persistence	1	1	-	-	$A_1 = 0.98$
AR(4) variable persistence	1	4	0.9	0.5	-
VAR(2) variable persistence	3	2	0.9	0.5	-

Omnibus exercise

Finally, to explore performance in very complex scenarios, we consider a single omnibus model incorporating a large number of unrelated series, Cauchy distributed innovations, and vector autoregressive transitions with strong persistence. The parameterization is given in Table 12.

TABLE 12. Changed parameter specifications for omnibus exercise

	$\%n_1$	$\%n_2$	$\%n_3$	r	p	ν	ξ	σ^2	u_γ
omnibus 1	20%	20%	60%	3	4	1	0.9	10	0.5

Evaluation criteria

In order to evaluate the performance of the estimators, we consider two criteria that are standard in the literature. The first is a multivariate trace R^2 statistic, defined to be

$$\hat{R}^2(F, \hat{F}) = \frac{tr(F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'F)}{tr(F'F)}$$

This statistic captures the extent to which the estimated factors span the same space as the true unobserved factors. Better performance is indicated by an R^2 statistic close to one.

The second statistic is the root mean squared forecast error from a one-step-ahead forecast.

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_{i,T+1|T} - y_{i,T+1})^2}{n}}$$

Better performance is indicated by a smaller $RMSE$ statistic.

Procedure

Each Monte Carlo exercise generates observed series and unobserved factors according to the model above, estimates the factors and parameters using each of the estimators, performs one out-of-sample forecast, and computes the R^2 and $RMSE$ statistics above. All exercises are performed separately for $T = 50, 100$ and $n = 5, 10, 25, 50, 100$; each exercise is replicated 100 times.

There are two differences to note in estimation for the PCA method. First, the data were standardized to have sample mean zero and sample standard deviation one prior to computing the principal components. Second, forecasts were not computed for this method. For this reason, the $RMSE$ results shown below do not have a PCA column.

For Gibbs sampling we must specify the number of burn-in draws and the number of draws from the posterior. The results reported below are based on the median value of the parameters from from 100 draws from the posterior, after 100 burn-in draws.[20]

Results

Results are reported in Table 13 for the case $T = 100$ and are based on median values across replications.[21] The first two columns identify the exercise and the next two columns give the value of the R^2 and $RMSE$ statistics for the 2-step estimator. The next four columns give the value of the R^2 statistic for the other estimators relative to the 2-step value, and the last three columns give the value of the $RMSE$ statistic for the other estimators relative to the 2-step value.

[20] These number of draws for both burn-in and from the posterior are quite small, due to computational constraints. Simulation of exercises with a larger number of draws are on-going, although preliminary results suggest that it does not substantially affect results.

[21] A similar table for the $T = 50$ case is reported in the Appendix, along with all other results.

For the columns reporting relative statistics, cells are lightly shaded if they represent any improvement over the 2-step estimator and are darkly shaded if they represent at least a 10% improvement. For all R^2 columns higher values imply better performance; for the $RMSE$ columns lower values imply better performance.

Comparisons across exercises are possible using the values reported for the 2-step estimator. For example, from these columns it is clear that the ability of the estimators to recover the true factors degrades substantially in the presence of outliers, relative to the baseline model.

Comparisons across estimators for each of the exercises are possible using the relative statistics. For example, it is easy to see that in the presence of outliers, the Gibbs sampler is generally able to recover more of the true factor space than any other estimator (of course, in absolute terms, it is still not doing a particularly good job).

The distribution of the statistics across the replications is also informative, and we employ scatterplots to present this information visually. A scatterplot for the baseline exercise is given in Fig. 8 for the case $T = 100$. To improve clarity, the combination EM and quasi-Newton estimator is not shown because it is nearly identical in all cases to the EM estimator.

TABLE 13. Evaluation statistics from Monte Carlo exercises

		R^2	$RMSE$	$R^2_{\cdot}/R^2_{\text{2-step}}$				$RMSE_{\cdot}/RMSE_{\text{2-step}}$		
	n	2-step		PCA	EM	EM,QN	GS	EM	EM,QN	GS
baseline:	5	0.83	2.48	0.81	1.00	1.00	1.00	0.96	0.96	0.95
AR(1)	10	0.93	2.38	0.84	1.00	1.00	0.99	1.03	1.03	1.02
	25	0.97	2.94	0.87	1.00	1.00	1.00	0.98	0.98	0.99
	50	0.98	3.03	0.92	1.01	1.01	1.01	0.98	0.98	1.00
	100	0.99	3.00	0.89	1.01	1.01	1.01	1.00	1.00	1.00

Continued on next page

TABLE 13. Evaluation statistics from Monte Carlo exercises

		R^2	$RMSE$	$R^2_{\cdot}/R^2_{\text{2-step}}$				$RMSE_{\cdot}/RMSE_{\text{2-step}}$		
	n	2-step		PCA	EM	EM,QN	GS	EM	EM,QN	GS
baseline: VAR(1)	5	0.59	4.18	0.80	1.00	1.00	1.13	1.01	1.01	0.98
	10	0.69	4.81	0.75	1.11	1.11	1.12	1.06	1.07	1.03
	25	0.89	4.96	0.80	1.02	1.02	1.03	1.01	1.01	1.04
	50	0.95	5.42	0.84	1.00	1.00	1.01	1.00	1.00	1.00
	100	0.97	5.33	0.84	1.00	1.00	1.01	1.00	1.00	1.01
outliers: cauchy	5	0.02	5.01	0.72	1.04	1.04	1.91	1.05	1.05	1.14
	10	0.02	6.20	1.08	1.00	0.87	1.69	1.04	1.04	1.12
	25	0.02	5.97	0.61	0.85	0.85	1.00	1.04	1.04	1.05
	50	0.02	7.73	0.76	0.89	0.89	0.97	0.96	0.96	1.04
	100	0.01	7.28	0.91	1.02	1.02	1.21	0.99	0.99	1.00
outliers: cauchy + removal	5	0.02	5.27	0.73	1.03	1.03	3.11	1.00	1.00	1.16
	10	0.03	7.82	0.59	1.29	1.29	1.39	0.94	0.94	1.06
	25	0.03	7.85	0.65	1.25	1.25	1.92	1.02	1.02	0.95
	50	0.02	9.06	0.84	0.99	0.99	2.28	0.97	0.97	0.99
	100	0.03	8.21	0.75	1.05	1.05	1.36	0.95	0.95	0.97
outliers: dof=2	5	0.12	2.79	0.57	1.54	1.54	2.51	1.02	1.02	1.07
	10	0.31	3.15	0.68	0.98	0.98	1.45	0.96	0.96	0.97
	25	0.44	4.08	0.80	1.30	1.30	1.33	1.04	1.04	1.01
	50	0.76	4.23	0.72	1.07	1.07	1.10	1.01	1.01	1.01
	100	0.76	4.40	0.60	1.08	1.08	1.10	1.02	1.03	1.01
outliers: dof=2 + removal	5	0.24	2.89	0.63	1.76	1.76	1.95	0.99	0.99	0.98
	10	0.56	3.77	0.74	1.10	1.10	1.14	1.00	1.00	1.05
	25	0.78	4.22	0.74	1.05	1.05	1.05	0.95	0.95	0.98
	50	0.89	4.65	0.78	1.03	1.03	1.02	1.02	1.02	0.98
	100	0.94	4.66	0.88	1.01	1.01	1.02	1.01	1.01	0.99
outliers: dof=10	5	0.84	2.24	0.80	1.01	1.01	1.01	0.95	0.95	0.98
	10	0.92	2.65	0.85	1.00	1.00	1.00	1.02	1.02	1.04
	25	0.96	3.19	0.93	1.00	1.00	1.00	1.02	1.02	1.01
	50	0.98	3.09	0.88	1.00	1.00	1.00	1.00	1.00	0.97
	100	0.98	3.30	0.93	1.01	1.01	1.01	1.02	1.02	0.99
unrelated: 20% 10x var	5	0.77	3.43	0.79	1.04	1.04	1.04	0.99	0.99	0.99
	10	0.90	4.12	0.84	1.01	1.01	0.99	0.97	0.97	0.97
	25	0.96	4.70	0.89	1.00	1.00	1.00	0.99	0.99	0.99
	50	0.98	4.73	0.92	1.01	1.01	1.01	0.99	0.99	1.00
	100	0.98	4.92	0.94	1.01	1.01	1.01	1.00	1.00	1.00

Continued on next page

TABLE 13. Evaluation statistics from Monte Carlo exercises

		R^2	$RMSE$	$R^2_{\cdot}/R^2_{\text{2-step}}$				$RMSE.\,/RMSE_{\text{2-step}}$		
	n	2-step		PCA	EM	EM,QN	GS	EM	EM,QN	GS
unrelated:	5	0.07	2.14	1.07	3.81	3.81	5.91	0.99	0.99	0.98
60% 1x var	10	0.54	2.70	0.63	1.48	1.48	1.49	1.01	1.01	1.02
	25	0.93	2.75	0.81	1.00	1.00	1.00	1.00	1.00	1.00
	50	0.96	2.95	0.86	1.00	1.00	1.00	0.99	0.99	1.01
	100	0.97	3.02	0.91	1.00	1.00	1.01	0.99	0.99	0.99
unrelated:	5	0.10	5.45	0.94	1.13	1.13	6.03	1.01	1.01	1.00
60% 10x var	10	0.41	6.79	0.63	1.34	1.34	1.98	0.95	0.95	1.02
	25	0.91	7.06	0.84	1.01	1.01	1.00	0.99	0.99	1.00
	50	0.95	7.43	0.87	1.00	1.00	1.00	1.00	1.00	1.00
	100	0.97	7.65	0.91	1.01	1.01	1.01	0.99	0.99	1.00
AR(1): high	5	0.72	4.91	0.39	1.20	1.20	1.20	0.96	0.96	0.94
persistence	10	0.87	5.14	0.52	1.05	1.05	1.05	0.98	0.98	0.95
	25	0.95	5.74	0.61	1.00	1.00	1.01	1.00	1.00	1.00
	50	0.98	6.45	0.64	0.99	0.99	1.00	0.99	0.99	0.98
	100	0.99	6.27	0.70	0.99	0.99	1.00	0.99	0.99	0.98
AR(1): low	5	0.84	1.14	0.99	0.92	0.92	0.93	0.98	0.98	0.97
persistence	10	0.93	1.41	0.98	0.99	0.99	0.98	1.00	1.00	0.99
	25	0.97	1.53	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	50	0.98	1.44	0.99	1.00	1.00	1.00	0.99	0.99	0.99
	100	0.99	1.59	0.99	1.00	1.00	1.00	1.02	1.02	1.02
AR(4):	5	0.08	3.55	0.25	0.93	0.93	0.90	1.01	0.98	1.02
variable	10	0.12	4.04	0.30	0.73	0.73	1.04	0.99	0.99	0.98
persistence	25	0.37	4.00	0.41	0.58	0.57	0.55	0.99	0.99	1.01
	50	0.17	5.09	0.42	0.74	0.74	1.59	0.98	0.98	0.99
	100	0.83	4.72	0.83	0.81	0.81	1.02	0.98	0.98	1.00
VAR(2):	5	0.53	4.71	0.75	0.74	0.71	0.99	0.96	0.98	0.90
variable	10	0.61	5.45	0.77	0.88	0.84	1.06	1.01	1.02	0.97
persistence	25	0.81	5.86	0.87	1.02	1.00	1.04	1.00	0.99	0.98
	50	0.87	6.36	0.86	0.97	0.97	1.01	0.99	0.99	0.99
	100	0.93	6.11	0.90	0.99	1.00	1.00	1.00	1.00	1.00
Omnibus: 1	5	0.13	54.13	0.21	0.95	0.97	0.89	1.04	1.04	1.03
	10	0.13	69.65	0.24	0.93	0.94	0.92	1.05	1.05	1.03
	25	0.12	66.81	0.21	1.01	1.01	1.09	1.00	1.04	1.03
	50	0.12	51.00	0.21	0.97	0.98	0.92	1.35	1.39	0.89

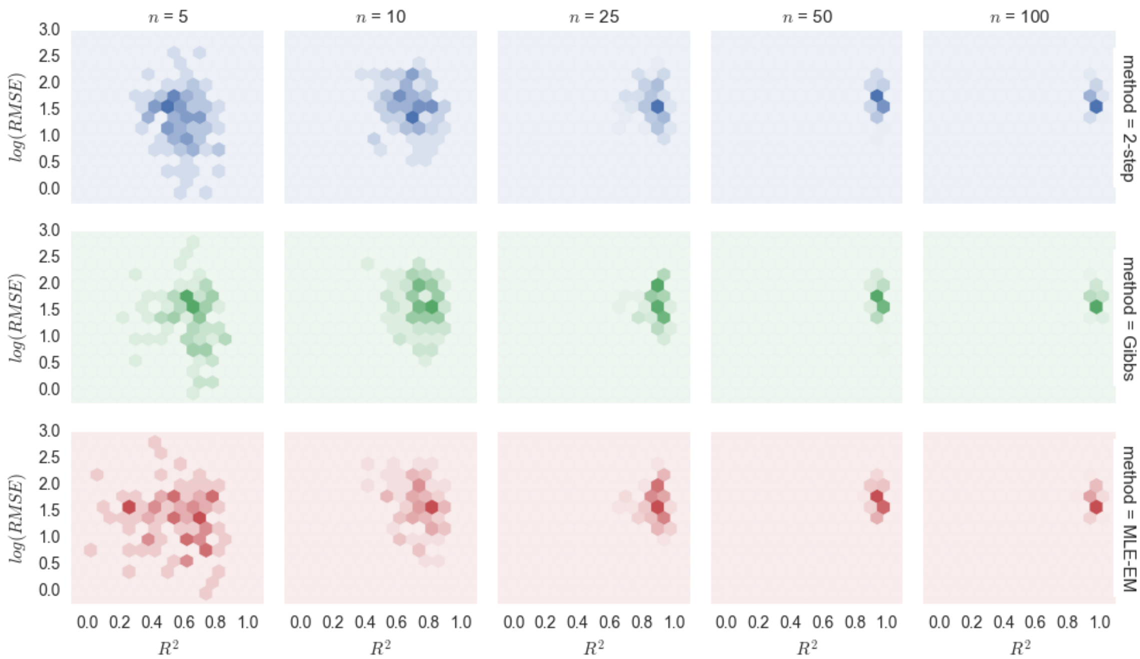

FIGURE 8. Hexagon-binned scatterplot of evaluation statistics for baseline: VAR(1)

Baseline

Results in the baseline model are qualitatively very similar to those of Doz et al. (2012) and generally reiterate a few well-known facts about dynamic factor models.

First, the estimators generally perform very well, recovering the majority of the true factor space. Second, in a well-specified model, the estimators making use of a parametric specification are more generally more efficient in factor factor estimation. Third, with the exception of the PCA, the performance of the estimators is very similar, and it is hard to draw definitive conclusions about the superiority of one estimator to another. Finally, as the number of "clean" observations increases, all estimators are better able to estimate the factors.

In particular, we note that the two estimators not previously considered in Monte Carlo studies, the combination EM and quasi-Newton estimator and the Gibbs sampler, share these well-known characteristics.

Considering the distribution of results across repetitions in Fig. 8, the evaluation statistics appear to be similarly dispersed for all the estimators. Increased sample size corresponds to reduced dispersion of both statistics for all estimators, with the reduction occurring at roughly the same rate.

Factor estimation

Focusing first specifically on the R^2 statistics, several results stand out. The first is that the Gibbs sampler tends to weakly outperform the other estimators, although this is by no means uniform. In most cases in which it does perform better, the difference is not dramatic. Nonetheless, if on the basis of these results one estimator had to be selected, it would be the Gibbs sampler. Counterbalanced against this is the fact that it is the most computationally intensive estimator: whereas the 2-step iterator requires only a single run of the KFS and the EM-algorithm usually requires less than 100 applications of the KFS, the Gibbs sampler requires as many applications as there are MCMC draws.[22] Since this is often thousands or tens of thousands in practice, the difference in computation time can be substantial, particularly for models with a large number of observed series. However, by employing the univariate or collapsed KFS approaches mentioned above, estimation by Gibbs sampler is still easily feasible on modern desktop or laptop computers.

[22] If a Metropolis-within-Gibbs approach were required, even more applications of the Kalman filter would be required.

On the other hand, the PCA estimator consistently offered the worst performance in recovering the true factors. This suggests that the only slightly more complicated two-step approach should be used in its place.

Next, the maximum likelihood estimators had relatively poor performance in exercises with more complicated factor transition dynamics. In the AR(4) and VAR(2) models, the MLE estimators were consistently worse than the two-step and Gibbs sampling approaches. For a more detailed look, the distribution of the statistics across repetitions is plotted in Fig. 9. Two groups of results are clear across all three estimators: a poorly estimated group in the upper left had corner of each graph and a well estimated group in the lower right hand corner. We conjecture that in the poorly estimated group, the MLE model converges to a local maximum that is suboptimal globally, whereas the Gibbs sampler and two-step estimators are better able to explore a wide range of models, resulting in a greater number of replicates with intermediate values of the statistics. The $n = 50$ case is particularly telling.[23]

It is also evident that the EM algorithm is able to converge quickly enough to the local maximum that adding subsequent quasi-Newton iterations does not noticeably improve results. In many cases, the convergence criterion for the quasi-Newton method (based on the norm of the score vector) was immediately satisfied by the parameters selected by the EM algorithm. Based on these results, and since computation of the analytic score adds an additional layer of complexity to estimation, it appears safe to use the EM algorithm by itself.[24]

---

[23] It would be interesting to investigate the effects of trying different starting parameters for MLE or making use of global optimization techniques. This is left for future work.

[24] However, in some more complicated dynamic factor settings, the maximization step of the EM algorithm no longer has a closed form solution and so the score vector is required anyway. See Jungbacker et al. (2011) for details.

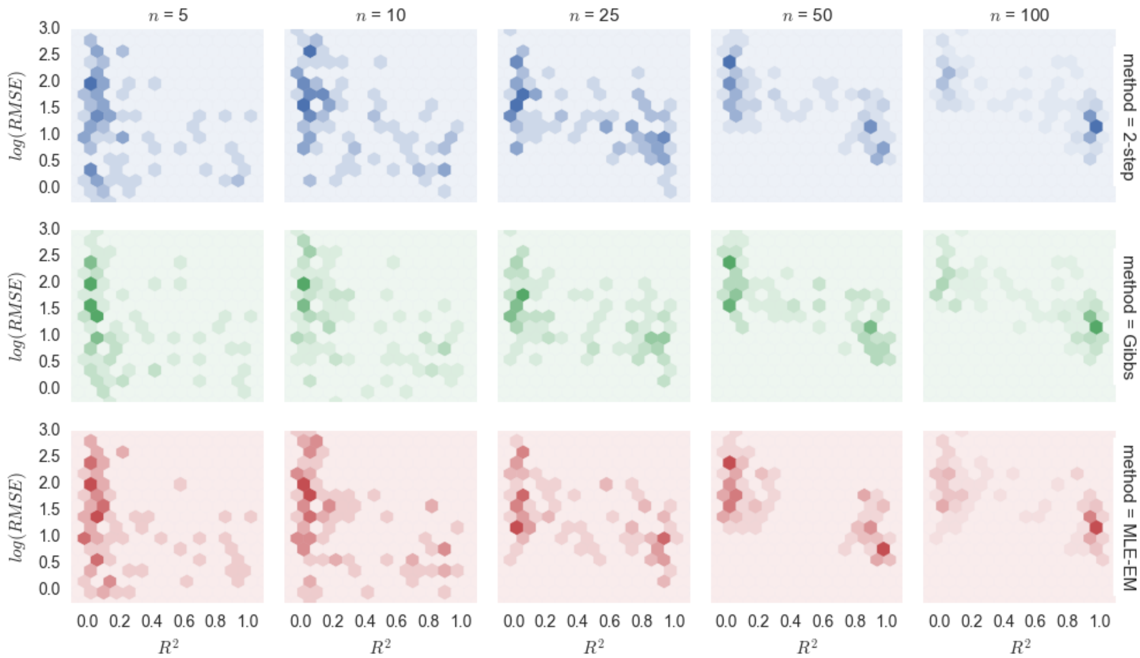

FIGURE 9. Hexagon-binned scatterplot of evaluation statistics for VAR(2): variable persistence

The third result is that the presence of outliers makes a substantial impact on the ability of all estimators to recover the true factor space, and that the removal of outliers can improve factor estimation, although this comes at the expense of forecasting performance. The estimators were essentially unable to estimate the factors with Cauchy disturbances regardless of the sample size (this is not necessarily unexpected since the first two moments of the Cauchy distribution do not exist and it is often described as a pathological distribution). In the two degrees of freedom case, all estimators performed better and were able to recover most of the factor space with a large enough number of observed series. Finally, the parametric estimators were almost always superior to the 2-step estimator, with the Gibbs sampler often performing much better.

Next, with enough data the presence of unrelated series appeared to have little effect on the performance of the estimators. However, for $n = 5, 10$ the parametric estimators were substantially better.

Finally, the results of the omnibus exercise suggest that the 2-step approach offers some protection in the case of extreme misspecification. Although it was still unable to recover much of the factor space, it generally outperformed all other estimators.

Forecasting

The evidence is much less clear about the preferred estimator from the one-step-ahead out-of-sample forecasting exercise. Each of the parametric estimators had superior performance in roughly half the cases, while the 2-step estimator was superior in the other half. Furthermore, the $RMSE$ statistics are much more similar across all models than were the R^2 statistics. All but a few lie within 5% of each other and most are within 3%, so even in cases when a given estimator was consistently preferred (for example parametric methods in the AR(1) high persistence exercise), the absolute differences were still quite small.

As one might expect, the $RMSE$ and R^2 statistics were negatively correlated overall, suggesting that better estimation of the factors corresponds to improved forecasting performance, although again this does not hold across all exercises. In the "outliers: dof=2" exercise, the parametric estimators had substantially higher R^2 statistics and yet largely performed worse at forecasting.

Conclusion

In this paper, we consider the finite-sample properties of non-parametric, classical, and Bayesian estimators of dynamic factor models. Two of our estimators

- maximum likelihood estimation (MLE) using quasi-Newton methods and Bayesian Gibbs sampling - have been previously used in empirical applications but have not been included in previous simulation studies, largely due to computational limitations. We also consider a more complex set of data generating processes than have previously been studied. Finally, in addition to evaluating estimators on the basis of summary statistics, we explore the distribution of evaluation statistics across the Monte Carlo repetitions.

First, we replicate previous results that find similarly good performance across all estimators in terms of recovering the underlying factor space when the estimating model is well-specified, and we show that these results extend to the two estimators not previously considered in Monte Carlo studies.

Across all scenarious and exercises, we find that the Gibbs sampling estimator appears to be the weakly superior method, although it does not uniformly dominate other methods and in absolute terms its improvements are often small. In selecting an estimator, these weak improvements must be balanced against its larger computational requirements. We further find that the addition of quasi-Newton MLE steps does not noticeably improve upon the more commonly used expectation maximization algorithm, suggesting that concerns about its slow theoretical convergence may not be relevant in practice.

Next, although the most basic non-parametric principal components method consistently performs the worst, its recent evolution as a hybrid “two-step” estimator is competitive with the parametric MLE and Gibbs sampling estimators. In particular, one-step-ahead out-of-sample forecasting performance is quite similar for all estimators, and R^2 summary statistics are similar when the sample size is large enough. These results combined with the modest computational requirements and ease of use of the two-step estimator make it an attractive option. In smaller samples, however, the parametric estimators are often better.

Our Monte Carlo exercises also consider model misspecifications that may be encountered in empirical work. We find that in the presence of outliers or when unrelated data series are included, the parametric estimators, and especially the Gibbs sampler, produce better estimates of the underlying factor space but do not necessarily produce improved forecasts. As the sample size grows, the two-step estimator becomes more competive. Across all estimators, we find that the common practice of eliminating outliers can improve the factor estimation but tends to degrade forecasting performance.

We find that in the presence of more complicated factor dynamics, the MLE estimators perform worse and conjecture based upon the distribution of statistics across Monte Calro repetitions that they may be more prone to getting stuck at local maxima. Finally, an omnibus exercise suggests that the two-step estimator may offer some protection in the case of extreme misspecification.

APPENDIX A

APPENDIX TO CHAPTER 2

Model

Constant Elasticity of Substitution Preferences

Definition: Consumption good

The composite consumption good is defined as a monotonic transformation of the generalized mean $\tilde{C}_{ht}$ as follows:

$$\tilde{C}_{ht} = \left[\frac{\sum_{i=1}^{I} \mu_i^{1-p} C_{hit}^{p}}{\sum_{i=1}^{I} \mu_i^{1-p}} \right]^{\frac{1}{p}}$$

$$C_{ht} = \left[\sum_{i=1}^{I} \mu_i^{1-p} C_{hit}^{p} \right]^{\frac{1}{p}} = \left[\sum_{i=1}^{I} \mu_i^{1-p} \right]^{\frac{1}{p}} \tilde{C}_{ht}$$

The exponent on the weight term is a normalization so that the resulting price index has the property that if every industry-level good has the same price, that price also is the index price. Furthermore, if all prices are the same then the derived demand for each industry-level good is just the fraction of the demand for the composite good weighted by the industry's size. Mathematically $C_{hit}^{d} = \mu_i C_{ht}$, and since we normalized the total measure of goods to one, $\mu_i \in [0, 1]$ for each industry i. This approach is the same as in Woodford (2003).

Definition: Industry-level good

The composite industry-level good is defined as a monotonic transformation of the generalized mean $\tilde{C}_{hit}$ as follows:

$$\tilde{C}_{hit} = \left[\frac{\int_{J_i} C_{hjt}^r dj}{\int_{J_i} 1 dj}\right]^{\frac{1}{r}} = \left[\int_{J_i} \mu_i^{-1} C_{hjt}^r dj\right]^{\frac{1}{r}}$$
$$C_{hit} = \left[\int_{J_i} \mu_i^{r-1} C_{hjt}^r\right]^{\frac{1}{r}} = \left[\mu_i^r \int_{J_i} \mu_i^{-1} C_{hjt}^r\right]^{\frac{1}{r}} = \mu_i \tilde{C}_{hit}$$

The exponent on the weight term is for the same normalizing purpose as above.

Demand: Constant Elasticity of Substitution Preferences

As in Dixit and Stiglitz (1977), we can use a multi-stage budgeting procedure to first solve for the demand for industry-level and intermediate goods' demand in terms of the consumers' total demand for the consumption good, and then solve their inter-temporal problem in terms only of the consumption good.

The first stage is itself split into two steps: (1) solve for industry-level demand in terms of total demand, and (2) solve for intermediate good demand in terms of industry-level demand.

Step 1: Industry-level demand The interpretation of the definition of the consumption good is as a *utility specification*. Thus solving for demand is the standard microeconomic constrained optimization problem

$$\max_{\{C_{hit}\}_{i=1}^I} u\left(\{C_i\}_{i=1}^I\right)$$

subject to $\sum_{i=1}^I C_{hit} P_i = W$, where P_{it} is the price of industry-level good i at time t and W is total wealth, and where the utility specification is the generalized mean, above:

$$u\left(\{C_i\}_{i=1}^{I}\right) = \tilde{C}_{ht} = \left[\frac{\sum_{i=1}^{I}\mu_i^{1-p}C_{hit}^{p}}{\sum_{i=1}^{I}\mu_i^{1-p}}\right]^{\frac{1}{p}}$$

The model used in the paper is a monotonic transformation of this specification, and it will yield equivalent demand specifications due the ordinal nature of utility functions.

This constrained optimization problem can be solved by forming a Lagrangian and taking first-order conditions. To ease notation, define $w_i \equiv \frac{\mu_i^{1-p}}{\sum_{i=1}^{I}\mu_i^{1-p}}$.

$$\mathcal{L} = \left[\sum_{i=1}^{I} w_i C_{hit}^{p}\right]^{\frac{1}{p}} - \lambda\left[\sum_{i=1}^{I} C_{hit}P_{it} - W\right]$$

Assuming an interior solution, the I first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial C_{hit}} = 0 = \frac{1}{p}\left[\sum_{i=1}^{I} w_i C_{hit}^{p}\right]^{\frac{1}{p}-1} w_i p C_{hit}^{p-1} - \lambda P_{it} = w_i u\left(\{C_i\}_{i=1}^{I}\right)^{1-p} C_{hit}^{p-1} - \lambda P_{it}$$

$$C_{hit} = \left(\frac{\lambda P_{it}}{w_i}\right)^{\frac{1}{p-1}} \tilde{C}_{ht}$$

This yields the demand for the industry-level good. The Lagrangian multiplier λ is the marginal value of relaxing the constraint, or the marginal value of wealth.

$$\begin{aligned}
\tilde{C}_{ht} &= \left[\sum_{i=1}^{I} w_i C_{hit}^{p}\right]^{\frac{1}{p}} = \left[\sum_{i=1}^{I} w_i \left(\left(\frac{\lambda P_{it}}{w_i}\right)^{\frac{1}{p-1}} \tilde{C}_{ht}\right)^{p}\right]^{\frac{1}{p}} \\
&= \tilde{C}_{ht}\lambda^{\frac{1}{p-1}}\left[\sum_{i=1}^{I} w_i^{\frac{1}{1-p}} P_{it}^{\frac{p}{p-1}}\right]^{\frac{1}{p}} \\
\frac{1}{\lambda} &= \left[\sum_{i=1}^{I} w_i^{\frac{1}{1-p}} P_{it}^{\frac{p}{p-1}}\right]^{\frac{p-1}{p}}
\end{aligned}$$

The price index is the price of the composite good $\tilde{C}_{ht}$, which is equivalently the price of increasing utility. This quantity is the inverse of the marginal value of wealth, so that

$$P_t \equiv \frac{1}{\lambda} = \left[\sum_{i=1}^{I} w_i^{\frac{1}{1-p}} P_{it}^{\frac{p}{p-1}}\right]^{\frac{p-1}{p}}$$

Notice that if all industry-level prices are the same, so that $P_{it} = P_{i't} = \bar{P}_t$, then:

$$\begin{aligned} P_t &= \left[\sum_{i=1}^{I} w_i^{\frac{1}{1-p}} \bar{P}_t^{\frac{p}{p-1}}\right]^{\frac{p-1}{p}} \\ &= \bar{P}_t \left[\sum_{i=1}^{I} w_i^{\frac{1}{1-p}}\right]^{\frac{p-1}{p}} \end{aligned}$$

Thus if we want to have the property that in this case $P_t = \bar{P}_t$, then we must have $\left[\sum_{i=1}^{I} w_i^{\frac{1}{1-p}}\right]^{\frac{p-1}{p}} = 1$. This does not hold for $\tilde{C}_{ht}$, but it does hold for the transformation C_{ht} since in that case $w_i \equiv \mu_i^{1-p}$ and then

$$\left[\sum_{i=1}^{I} w_i^{\frac{1}{1-p}}\right]^{\frac{p-1}{p}} = \left[\sum_{i=1}^{I} \mu_i^{\frac{1-p}{1-p}}\right]^{\frac{p-1}{p}} = 1^{\frac{p-1}{p}} = 1$$

Finally we can rewrite industry-level demand

$$C_{hit} = w_i^{\frac{1}{1-p}} \left(\frac{P_{it}}{P_t}\right)^{\frac{1}{p-1}} \tilde{C}_{ht}$$

If we use the transformed C_{ht}, then this reduces to:

$$C_{hit} = \mu_i^{\frac{1-p}{1-p}} \left(\frac{P_{it}}{P_t}\right)^{\frac{1}{p-1}} C_{ht}$$

Collecting the final demand function and price index for the transformed C_{ht}, we have

$$C_{hit} = \mu_i \left(\frac{P_{it}}{P_t} \right)^{\frac{1}{p-1}} C_{ht}$$

$$P_t = \left[\sum_{i=1}^{I} \mu_i P_{it}^{\frac{p}{p-1}} \right]^{\frac{p-1}{p}}$$

Step 2: Intermediate good demand Following similar steps as above, the final demand function and price index are given by

$$C_{hjt} = \frac{1}{\mu_i} \left(\frac{P_{jt}}{P_{it}} \right)^{\frac{1}{r-1}} C_{hit}$$

$$P_{it} = \left[\int_{J_i} \frac{1}{\mu_i} P_{jt}^{\frac{r}{r-1}} dj \right]^{\frac{r-1}{r}}$$

And the CES demand function for intermediate goods in terms of the consumption good is

$$\begin{aligned} C_{hjt} &= \frac{1}{\mu_i} \left(\frac{P_{jt}}{P_{it}} \right)^{\frac{1}{r-1}} \mu_i \left(\frac{P_{it}}{P_t} \right)^{\frac{1}{p-1}} C_{ht} \\ &= P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{1}{1-r} + \frac{1}{p-1}} P_t^{\frac{1}{1-p}} C_{ht} \end{aligned}$$

Budget Contraints

In each period, households purchase (1) consumption goods, and (2) invest in risk-free bonds. They receive income from (1) wages, (2) a share of intermediate goods firm profits, and (3) investment income from bonds purchased in the previous period.

Assume that all households are endowed with equal ownership shares in each of the intermediate goods firms. Then each household's share of the profits can be denoted π_{jt}.

Bonds are indexed by time period in which they mature so that B_t refers to bonds purchased in time $t-1$ that yield income in period t. The gross nominal rate of return on a bond purchased in period $t-1$ is denoted R_t. The bonds are riskless, so that R_t is known in period $t-1$.

The nominal flow budget constraint is

$$\int_J P_{jt}C_{hjt}dj + B_{ht+1} \leq \int_J \theta_{hj}\pi_{jt}dj + \int_J W_{jt}n_{hjt}dj + R_t B_{ht}$$

Plugging the CES demand functions derived above into the consumption spending portion of the budgent constraint yiels

$$\begin{aligned}
\int_J P_{jt}P_{jt}^{\frac{1}{r-1}}P_{it}^{\frac{1}{1-r}+\frac{1}{p-1}}P_t^{\frac{1}{1-p}}C_{ht}dj &= P_t^{\frac{1}{1-p}}C_{ht}\int_J P_{jt}^{\frac{r}{r-1}}P_{it}^{\frac{1}{1-r}+\frac{1}{p-1}}dj \\
&= P_t^{\frac{1}{1-p}}C_{ht}\sum_{i=1}^{I}\mu_i P_{it}^{\frac{r}{r-1}}P_{it}^{\frac{1}{1-r}} \\
&= P_t^{\frac{1}{1-p}}C_{ht}\sum_{i=1}^{I}\mu_i P_{it}^{\frac{p}{p-1}} \\
&= P_t^{\frac{1}{1-p}}C_{ht}P_t^{\frac{p}{p-1}} \\
&= P_t C_{ht}
\end{aligned}$$

Using this, the nominal flow budget constraint can be rewritten

$$P_t C_{ht} + B_{ht+1} \leq \int_J \theta_{hj} \pi_{jt} dj + \int_J W_{jt} n_{hjt} dj + R_t B_{ht}$$

Optimal Behavior

Optimal Household Behavior

Sequential Problem

The representative household's problem is

$$\max_{\{C_t\}_{t\geq 0}\{n_{jt}\}_{j\in J, t\geq 0}} = E_0 \sum_{t=0}^{\infty} \beta^t \left[u(C_t) - \int_J v(n_{jt}) dj \right]$$

subject to the nominal budget constraint

$$P_t C_t + B_{t+1} \leq \int_J \pi_{jt} dj + \int_J W_{jt} n_{jt} dj + R_t B_t$$

Define wealth at time t as

$$A_t = \int_J \pi_{jt} dj + \int_J W_{jt} n_{jt} dj + R_t B_t$$

Notice that given wealth and the household's consumption choice, bond holdings are determined by $B_{t+1} = A_t - P_t C_t$.

Bellman system

The solution to the sequential problem is equivalent to the solution to the following functional equation

$$V(A) = \max_{C,\{n_j\}_{j\in J}} \left\{ u(C) - \int_J v(n_j)dj + \beta E\left[V(A')\right] \right\}$$

subject to

$$\begin{aligned} A' &= \int_J \pi_j' dj + \int_J W_j' n_j' dj + R'B' \\ &= \int_J \pi_j' dj + \int_J W_j' n_j' dj + R'(A - PC) \end{aligned}$$

First-order Conditions

$$\begin{aligned} 0 &= \frac{\partial V(A)}{\partial C} = u'(C) + \beta E\left[V'(A')\right](-P)R' \\ 0 &= \frac{\partial V(A)}{\partial n_j} = -v'(n_j) + \beta V'(A')W_j R' \end{aligned}$$

Envelope Condition

$$V'(A) = \beta V'(A')R'$$

Euler Equation Combining the first-order condition for consumption and the envelope condition yields

$$V'(A) = \frac{u'(C)}{P}$$

which can then be forwarded and plugged back into the first-order condition for consumption to give the household's Euler equation governing intertemporal consumption tradeoffs

$$u'(C_t) = \beta E_t \left[R_{t+1} \frac{P_t}{P_{t+1}} u'(C_{t+1}) \right]$$

Static First-order Condition Then from the first-order condition for labor we get

$$v'(n_{jt}) = \frac{W_{jt}}{P_t} u'(C_t)$$

Optimal Price Setting

Perfect Information

Firms face the problem

$$\max_{P_{jt}} u'(C_t) \left(P_{jt} - \frac{W_{jt}}{\varphi_{it}} \right) Y_{jt}$$

which can be rewritten using the CES demand function as

$$\max_{P_{jt}} u'(C_t) \left(P_{jt} - \frac{W_{jt}}{\varphi_{it}} \right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{1}{1-r} + \frac{1}{p-1}} P_t^{\frac{1}{1-p}} C_t$$

Their first-order condition is:

$$
\begin{aligned}
0 &= u'(C_t)\left[\left(1+\frac{1}{r-1}\right)P_{jt}^{\frac{1}{r-1}} - \left(\frac{1}{r-1}\frac{W_{jt}}{\varphi_{it}}\right)P_{jt}^{\frac{1}{r-1}-1}\right]P_{it}^{\frac{1}{1-r}+\frac{1}{p-1}}P_t^{\frac{1}{1-p}}C_t \\
rP_{jt}^{\frac{1}{r-1}} &= \frac{W_{jt}}{\varphi_{it}}P_{jt}^{\frac{1}{r-1}-1} \\
P_{jt} &= \frac{1}{r}\frac{W_{jt}}{\varphi_{it}}
\end{aligned}
$$

This is the standard result that monopolists set price as a markup over marginal costs.

Proceed by substituting out wages using the household's static first-order condition and using (1) the goods market clearing condition, (2) the production function, and (3) the demand function for the intermediate good

$$
\begin{aligned}
P_{jt} &= \frac{1}{r}\frac{1}{\varphi_{it}}\left[P_t\frac{n^{\varepsilon}}{C_t^{-\sigma}}\right] \\
&= \frac{1}{r}\frac{1}{\varphi_{it}}\left[P_tY_t^{\sigma}\left(\frac{Y_{jt}}{\varphi_{it}}\right)^{\varepsilon}\right] \\
&= \frac{1}{r}\left(\frac{1}{\varphi_{it}}\right)^{1+\varepsilon}\left[P_tY_t^{\sigma}\left(P_{jt}^{\frac{1}{r-1}}P_{it}^{\frac{1}{1-r}+\frac{1}{p-1}}P_t^{\frac{1}{1-p}}Y_t\right)^{\varepsilon}\right]
\end{aligned}
$$

Since productivity shocks are industry-level and they represent the only difference between firms, we can now apply symmetry between all firms within a given industry to note that $P_{jt} = P_{it}$.

$$
\begin{aligned}
P_{jt} &= \frac{1}{r}\left(\frac{1}{\varphi_{it}}\right)^{1+\varepsilon}\left[P_{jt}^{\frac{\varepsilon}{p-1}}P_t^{\frac{1-p+\varepsilon}{1-p}}Y_t^{\sigma+\varepsilon}\right] \\
P_{jt}^{\frac{1-p+\varepsilon}{1-p}} &= \left[\frac{1}{r}\left(\frac{1}{\varphi_{it}}\right)^{1+\varepsilon}Y_t^{\sigma+\varepsilon}\right]P_t^{\frac{1-p+\varepsilon}{1-p}} \\
P_{jt} &= \left[\frac{1}{r}\left(\frac{1}{\varphi_{it}}\right)^{1+\varepsilon}Y_t^{\sigma+\varepsilon}\right]^{\frac{1-p}{1-p+\varepsilon}}P_t^{\frac{1-p+\varepsilon}{1-p+\varepsilon}}
\end{aligned}
$$

Defining $\alpha \equiv \frac{1-p}{1-p+\varepsilon} = \frac{1}{1+\rho\varepsilon}$ we arrive at the final perfect information pricing equation

$$P_{jt}^{\diamond} = \left[\frac{1}{r} \left(\frac{1}{\varphi_{it}} \right)^{1+\varepsilon} Y_t^{\sigma+\varepsilon} \right]^{\alpha} P_t$$

It will also be convenient to have this expression with variables in log-form, where lowercase variables denote logs of uppercase variables

$$p_{jt}^{\diamond} = \alpha \log \frac{1}{r} - \alpha(1+\varepsilon)\phi_{it} + \alpha(\sigma+\varepsilon)y_t + p_t$$

To expose strategic complementarities define $\zeta = \alpha(\sigma+\varepsilon)$, and for notational convenience define $\gamma = \alpha(1+\varepsilon)$. Recall also that $q_t = p_t + y_t$. Then the firms' perfect information pricing rule is

$$p_{jt}^{\diamond} = \alpha \log \frac{1}{r} - \gamma\phi_{it} + \zeta q_t + (1-\zeta)p_t$$

To aid interpretation of the imperfect information pricing rule we express the pricing-rule in proportional deviation from common price form below by defining $\tilde{x}_t \equiv \frac{(X_t - \bar{X})}{\bar{X}} \approx \log\left(\frac{X_t}{\bar{X}}\right) = x_t - \bar{x}$

$$p_{jt}^{\diamond} - \bar{p}_{jt} = \left(\alpha \log \frac{1}{r} - \gamma\phi_{it} + \zeta q_t + (1-\zeta)p_t \right) - \left(\alpha \log \frac{1}{r} - \gamma\bar{\phi}_{it} + \zeta\bar{q}_t + (1-\zeta)\bar{p} \right)$$

which reduces to

$$\tilde{p}_{jt}^{\diamond} = -\gamma\tilde{\phi}_{it} + \zeta\tilde{q}_t + (1-\zeta)\tilde{p}_t$$

Imperfect Information

This section describes optimal firm behavior under imperfect information using results based on log approximations; it is just a summary of the approximations derived in detail in *Log Approximations*.

The generic problem facing an intermediate goods firm is given above in (A.2). In equilibrium, the objective can be written as a function only of prices, shocks, and aggregate output (this is because wages can be substituted out as a function of these variables, using the households' static first-order condition)

$$\Pi_{jt}(P_{jt}, P_{it}, P_t, Y_t, \varphi_t)$$

Given this, the firm's problem can be expressed as

$$\max_{P_{jt}} \Pi_{jt}$$

We proceed by taking a log-quadratic approximation to Π_{jt} around the perfect-information equilibrium,

$$\begin{aligned}\tilde{\Pi}_{jt} =& \Pi_1 \bar{P}\tilde{p}_{jt} + \frac{\Pi_{11}}{2!}\bar{P}^2\tilde{p}_{jt}^2 + \Pi_{12}\bar{P}^2\tilde{p}_{jt}E_{jt}\tilde{p}_{it} + \Pi_{13}\bar{P}^2\tilde{p}_{jt}E_{jt}\tilde{p}_t + \Pi_{14}\bar{P}\bar{Q}\tilde{p}_{jt}E_{jt}\tilde{y}_t \\ &+ \Pi_{15}\bar{P}\bar{\varphi}_{it}\tilde{p}_{jt}E_{jt}\tilde{\phi}_{it} \\ &+ \text{other terms}\end{aligned}$$

where Π_1 is the partial derivative of profit with respect to the first argument (P_{jt}) and the $\Pi_{1\cdot}$ coefficients are the partial derivatives of Π_1, all evaluated at the perfect-information equilibrium values, and $\bar{P}$ is the perfect-information equilibrium price. The "other terms" are all other terms in the second-order approximation that do not depend on $\tilde{p}_{jt}$, which are irrelevent here since they will not affect the firm's pricing decision.

The problem faced by firm j can now be written

$$\max_{P_{jt}} \tilde{\Pi}_{jt}$$

And the solution is characterized by the first-order condition

$$\begin{aligned}\frac{\partial\tilde{\Pi}}{\partial\tilde{p}_{jt}} = 0 = &\Pi_1\bar{P} + \Pi_{11}\bar{P}^2\tilde{p}_{jt} + \Pi_{12}\bar{P}^2E_{jt}\tilde{p}_{it} + \Pi_{13}\bar{P}^2E_{jt}\tilde{p}_t \\ &+ \Pi_{14}\bar{P}\bar{Q}E_{jt}\tilde{Q}_t + \Pi_{15}\bar{P}\bar{\varphi}_{it}\tilde{E}_{jt}\tilde{\phi}_{it}\end{aligned}$$

which reduces to

$$\begin{aligned}\tilde{p}^*_{jt} &= -\gamma E_{jt}\tilde{\phi}_{it} + \zeta E_{jt}\tilde{q}_t + (1-\zeta)E_{jt}\tilde{p}_t \\ &= E_{jt}\tilde{p}^{\diamond}_{jt}\end{aligned}$$

Log Approximations

Log-linear approximation to the aggregate price index

For results that follow, we will require a log-linear approximation to the price index P_t. Recall from *Constant Elasticity of Substitution Preferences* that P_t is derived as the (minimum) cost of purchasing one unit of the consumption good and is defined to be

$$P_t = \left[\sum_{i=1}^{I}\mu_i P_{it}^{\frac{p}{p-1}}\right]^{\frac{p-1}{p}}$$

We take the log-linear approximation around the where all prices are the same $P_{jt} = P_{it} = P_t \equiv \bar{P}$. Define $\tilde{p}_t \equiv \frac{(P_t - \bar{P})}{\bar{P}} \approx \log\left(\frac{P_t}{\bar{P}}\right) = p_t - \bar{p}$ so that $\tilde{p}_t$ is the aggregate price in proportional deviation-from-steady-state form.

$$\begin{aligned}
\bar{P} + (1)(P_t - \bar{P}) &= \bar{P} + \sum_{i=1}^{I} \frac{p-1}{p} P_t^{\frac{1}{1-p}} \mu_i P_{it}^{\frac{1}{p-1}} \frac{p}{p-1}(P_{it} - \bar{P}) \\
(P_t - \bar{P}) &= \sum_{i=1}^{I} \mu_i (P_{it} - \bar{P}) \\
\bar{P}\tilde{p}_t &= \sum_{i=1}^{I} \mu_i \bar{P}_i \tilde{p}_{it}
\end{aligned}$$

Thus the log-approximation aggregate price is described by

$$\tilde{p}_t = \sum_{i=1}^{I} \mu_i \tilde{p}_{it}$$

Log-quadratic approximation to an intermediate good firm's profit function

Recall from *Optimal Price Setting* that the problem faced by firm j can be written

$$\max_{P_{jt}} \Pi_{jt}(P_{jt}, P_{it}, P_t, Y_t, \varphi_{it})$$

A second-order approximation to this objective function around the perfect information non-stochastic equilibrium is given by

$$\begin{aligned}\bar{\Pi} + (1)(\Pi_{jt} - \bar{\Pi}) =& \bar{\Pi} + \Pi_1(P_{jt} - \bar{P}) + \frac{\Pi_{11}}{2!}(P_{jt} - \bar{P})^2 + \Pi_{12}(P_{jt} - \bar{P})(P_{it} - \bar{P}) \\ &+ \Pi_{13}(P_{jt} - \bar{P})E_{jt}(P_t - \bar{P}) + \Pi_{14}(P_{jt} - \bar{P})E_{jt}(Y_t - \bar{Y}) \\ &+ \Pi_{15}(P_{jt} - \bar{P})E_{jt}(\varphi_{it} - \bar{\varphi}_t) \\ &+ \text{other terms}\end{aligned}$$

where $\bar{P}$ denotes the price at which $P_{jt} = P_{it} = P_t \equiv \bar{P}$ and $\bar{Q}_t$ and $\bar{\varphi}_t$ denote the means of the processes. Π_1 is the partial derivative of profit with respect to the first argument Π_{1*} represent second partial derivatives, all evaluated at the point at which all prices are the same. The term "other terms" collects all terms that do not depend on $\tilde{p}_{jt}$ (irrelevent for our purposes since they do not affect the firm's pricing decision). The forms of these partial derivatives are derived below.

In log-deviation form the objective function is

$$\begin{aligned}\tilde{\Pi}(\tilde{p}_{jt}, \tilde{p}_{it}, \tilde{p}_t, \tilde{y}_t, \tilde{\phi}_{it}) =& \Pi_1\bar{P}\tilde{p}_{jt} + \frac{\Pi_{11}}{2!}\bar{P}^2\tilde{p}_{jt}^2 + \Pi_{12}\bar{P}^2\tilde{p}_{jt}\tilde{p}_{it} + \Pi_{13}\bar{P}^2\tilde{p}_{jt}E_{jt}\tilde{p}_t \\ &+ \Pi_{14}\bar{P}\bar{Y}\tilde{p}_{jt}E_{jt}\tilde{y}_t + \Pi_{15}\bar{P}\bar{\varphi}_t\tilde{p}_{jt}E_{jt}\tilde{\phi}_{it} \\ &+ \text{other terms}\end{aligned}$$

Derivatives

Below we calculate the first and second partial derivatives used in the log-quadratic approximation, above. The second partial derivatives are all first with respect to P_{jt} and then second with respect to the given variable. Evaluation of derivatives will be around the perfect information non-stochastic equilibrium (see *Perfect Information*) in which all prices are the same (we will use that the means of the shocks have been defined to be identical, see *Stochastic processes*, to guarantee the last condition).

Before calculating the derivatives, simplify the objective function by applying market clearing and household optimization so that it can be written

$$\begin{aligned}
\Pi_{jt} &= E_{jt}\left[U'(Y_t)\left(P_{jt}-\frac{W_{jt}}{\varphi_{it}}\right)Y_{jt}\right] \\
&= E_{jt}\left[Y_t^{-\sigma}\left(P_{jt}-\frac{W_{jt}}{\varphi_{it}}\right)\left(\frac{1}{\mu_i}\right)\left(\frac{P_{jt}}{P_{it}}\right)^{\frac{1}{r-1}}\mu_i\left(\frac{P_{it}}{P_t}\right)^{\frac{1}{p-1}}Y_t\right] \\
&= E_{jt}\left[\left(P_{jt}-\frac{W_{jt}}{\varphi_{it}}\right)P_{jt}^{\frac{1}{r-1}}P_{it}^{\frac{r-p}{(p-1)(r-1)}}P_t^{\frac{1}{1-p}}Y_t^{1-\sigma}\right] \\
&= E_{jt}\left[P_{it}^{\frac{r-p}{(p-1)(r-1)}}P_t^{\frac{1}{1-p}}Y_t^{1-\sigma}\left(P_{jt}^{1+\frac{1}{r-1}}-P_{jt}^{\frac{1}{r-1}}\frac{W_{jt}}{\varphi_{it}}\right)\right]
\end{aligned}$$

Since factor markets are perfectly competitive, so that firms are wage-takers, the wage cannot yet be substituted out.

First derivative, with respect to P_{jt}

$$\begin{aligned}
\frac{\partial \Pi_{jt}}{\partial P_{jt}} = E_{jt}[&\left(\frac{1}{r-1}\right)P_{jt}^{\frac{1}{r-1}}P_{it}^{\frac{r-p}{(p-1)(r-1)}}P_t^{\frac{1}{1-p}}Y_t^{1-\sigma} \\
&-\left(1+\frac{1}{r-1}\right)P_{jt}^{\frac{1}{r-1}-1}P_{it}^{\frac{r-p}{(p-1)(r-1)}}P_t^{\frac{1}{1-p}}Y_t^{1-\sigma}\frac{W_{jt}}{\varphi_{it}}]
\end{aligned}$$

After taking the derivative, wages can be substituted out using the firm's static first-order condition

$$
\begin{aligned}
\frac{\partial \Pi_{jt}}{\partial P_{jt}} &= E_{jt}\left(\frac{r}{r-1}\right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma} \\
&- E_{jt}\left(\frac{1}{r-1}\right) \frac{1}{\varphi_{it}} P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma} \left(P_t \frac{n_{jt}^{\varepsilon}}{Y_t^{-\sigma}}\right) \\
&= E_{jt}\left(\frac{r}{r-1}\right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma} \\
&- E_{jt}\left(\frac{1}{r-1}\right) \frac{1}{\varphi_{it}^{1+\varepsilon}} P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{2-p}{1-p}} Y_t \left[Y_{jt}\right]^{\varepsilon} \\
&= E_{jt}\left(\frac{r}{r-1}\right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma} \\
&- E_{jt}\left(\frac{1}{r-1}\right) \frac{1}{\varphi_{it}^{1+\varepsilon}} P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{2-p}{1-p}} Y_t \\
&\times \left[\left(\frac{1}{\mu_i}\right)\left(\frac{P_{jt}}{P_{it}}\right)^{\frac{1}{r-1}} \mu_i \left(\frac{P_{it}}{P_t}\right)^{\frac{1}{p-1}} Y_t\right]^{\varepsilon}
\end{aligned}
$$

This finally yields the first derivative of the log-quadratic approximation to the profit function with respect to price

$$
\begin{aligned}
\frac{\partial \Pi_{jt}}{\partial P_{jt}} =& E_{jt}\left(\frac{r}{r-1}\right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma} \\
&- E_{jt}\left(\frac{1}{r-1}\right) \frac{1}{\varphi_{it}^{1+\varepsilon}} P_{jt}^{\frac{1+\varepsilon}{r-1}-1} P_{it}^{(1+\varepsilon)\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{2-p+\varepsilon}{1-p}} Y_t^{1+\varepsilon}
\end{aligned}
$$

Evaluation at the point where all prices are the same implies $P_{jt} = P_{it} = P_t \equiv \bar{P}$. At this point we also have $\bar{\varphi}_{it} \equiv \bar{\varphi}_t$ for $i = 1, \ldots, I$. Then the exponent on the price on the left hand side is

$$
\begin{aligned}
\frac{1}{r-1} + \frac{r-p}{(p-1)(r-1)} - \frac{1}{p-1} &= \frac{(p-1)+(r-p)-(r-1)}{(p-1)(r-1)} \\
&= 0
\end{aligned}
$$

and the exponent on price on the right hand side is:

$$\frac{1+\varepsilon}{r-1}+(1+\varepsilon)\frac{r-p}{(p-1)(r-1)}-\frac{1+\varepsilon}{p-1}=(1+\varepsilon)\frac{(p-1)+(r-p)-(r-1)}{(p-1)(r-1)}$$
$$=0$$

this leads to

$$\Pi_1 \equiv \left.\frac{\partial \Pi_{jt}}{\partial P_{jt}}\right|_{\bar{P},\bar{Y},\bar{\varphi}_t}=\left(\frac{r}{r-1}\right)\bar{Y}^{1-\sigma}-\left(\frac{1}{r-1}\right)\left(\frac{1}{\bar{\varphi}_t}\right)^{1+\varepsilon}\bar{Y}^{1+\varepsilon}$$

Recall that $\bar{Y}=Y_t^n=r^{\frac{\alpha}{\zeta}}\left(\frac{1}{\bar{\varphi}_t}\right)^{-\frac{\gamma}{\zeta}}$ when shocks have a common mean (see *Perfect Information*). Then

$$\begin{aligned}
\Pi_1 \equiv \left.\frac{\partial \Pi_{jt}}{\partial P_{jt}}\right|_{\bar{P},\bar{Y},\bar{\varphi}_t} &= \left(\frac{r}{r-1}\right) r^{\frac{1-\sigma}{\sigma+\varepsilon}}\left(\frac{1}{\bar{\varphi}_t}\right)^{-(1-\sigma)\frac{1+\varepsilon}{\sigma+\varepsilon}} \\
&- \left(\frac{1}{r-1}\right)\left(\frac{1}{\bar{\varphi}_t}\right)^{1+\varepsilon}\left(\frac{1}{\bar{\varphi}_t}\right)^{-(1+\varepsilon)\frac{1+\varepsilon}{\sigma+\varepsilon}} r^{\frac{1+\varepsilon}{\sigma+\varepsilon}} \\
&= \left(\frac{1}{r-1}\right) r^{\frac{1-\sigma}{\sigma+\varepsilon}+1}\left(\frac{1}{\bar{\varphi}_t}\right)^{-(1-\sigma)\frac{1+\varepsilon}{\sigma+\varepsilon}} \\
&- \left(\frac{1}{r-1}\right)\left(\frac{1}{\bar{\varphi}_t}\right)^{1+\varepsilon-(1+\varepsilon)\frac{1+\varepsilon}{\sigma+\varepsilon}} r^{\frac{1+\varepsilon}{\sigma+\varepsilon}} \\
&= \left(\frac{1}{r-1}\right) r^{\frac{1+\varepsilon}{\sigma+\varepsilon}}\left(\frac{1}{\bar{\varphi}_t}\right)^{-(1-\sigma)\frac{1+\varepsilon}{\sigma+\varepsilon}} \\
&- \left(\frac{1}{r-1}\right)\left(\frac{1}{\bar{\varphi}_t}\right)^{1+\varepsilon-(1+\varepsilon)r^{\frac{1+\varepsilon}{\sigma+\varepsilon}}\frac{1+\varepsilon}{\sigma+\varepsilon}} \\
&= 0
\end{aligned}$$

where the last equality can be found by noting that the right-hand side exponent on $1/\varphi$ is the same as the left-hand side exponent:

$$\begin{aligned}(1+\varepsilon)-(1+\varepsilon)\frac{1+\varepsilon}{\sigma+\varepsilon} &= (1+\varepsilon)\left[1-\frac{1+\varepsilon}{\sigma+\varepsilon}\right]\\ &= (1+\varepsilon)\left[\frac{\sigma+\varepsilon-1-\varepsilon}{\sigma+\varepsilon}\right]\\ &= (1+\varepsilon)\left[\frac{\sigma-1}{\sigma+\varepsilon}\right]\\ &= -(1+\varepsilon)\left[\frac{1-\sigma}{\sigma+\varepsilon}\right] = -(1-\sigma)\left[\frac{1+\varepsilon}{\sigma+\varepsilon}\right]\end{aligned}$$

Second derivatives Note that

$$\begin{aligned}\bar{W} &= \frac{1}{\bar{\varphi}_t^{\varepsilon}}\bar{P}\bar{Y}^{\sigma}\left[\left(\frac{1}{\mu_i}\right)\left(\frac{\bar{P}}{\bar{\bar{P}}}\right)^{\frac{1}{r-1}}\mu_i\left(\frac{\bar{P}}{\bar{\bar{P}}}\right)^{\frac{1}{p-1}}\bar{Y}\right]^{\varepsilon}\\ &= \frac{1}{\bar{\varphi}_t^{\varepsilon}}\bar{P}\bar{Y}^{\sigma+\varepsilon}\\ &= r\bar{\varphi}_t\bar{P}\end{aligned}$$

Second derivative, with respect to P_{jt}

$$\begin{aligned}\frac{\partial^2\Pi_{jt}}{\partial P_{jt}^2} &= \frac{\partial}{\partial P_{jt}}E_{jt}\left(\frac{r}{r-1}\right)P_{jt}^{\frac{1}{r-1}}P_{it}^{\frac{r-p}{(p-1)(r-1)}}P_t^{\frac{1}{1-p}}Y_t^{1-\sigma}\\ &- \left(\frac{1}{r-1}\right)P_{jt}^{\frac{1}{r-1}-1}P_{it}^{\frac{r-p}{(p-1)(r-1)}}P_t^{\frac{1}{1-p}}Y_t^{1+\varepsilon}\frac{1}{\varphi_{it}^{1+\varepsilon}}P_t\left[\left(\frac{P_{jt}}{P_{it}}\right)^{\frac{1}{r-1}}\left(\frac{P_{it}}{P_t}\right)^{\frac{1}{p-1}}\right]^{\varepsilon}\\ &= E_{jt}\left(\frac{r}{r-1}\right)\left(\frac{1}{r-1}\right)P_{jt}^{\frac{1+\varepsilon}{r-1}-1}P_{it}^{\frac{r-p}{(p-1)(r-1)}}P_t^{\frac{1}{1-p}}Y_t^{1-\sigma}\\ &- \left(\frac{1}{r-1}\right)\left(\frac{1}{r-1}-1\right)P_{jt}^{\frac{1}{r-1}-2}P_{it}^{\frac{r-p}{(p-1)(r-1)}}P_t^{\frac{1}{1-p}}Y_t^{1-\sigma}\frac{W_{jt}}{\varphi_{it}}\end{aligned}$$

$$
\begin{aligned}
\Pi_{11} \equiv \left.\frac{\partial^2 \Pi_{jt}}{\partial P_{jt}^2}\right|_{\bar{P},\bar{Y},\bar{\varphi}_t} &= \left(\frac{r}{r-1}\right)\left(\frac{1}{r-1}\right)\bar{P}^{-1}\bar{Y}^{1-\sigma} \\
&\quad - \left(\frac{1}{r-1}\right)\left(\frac{2-r+\varepsilon}{r-1}\right)\bar{P}^{-2}\bar{Y}^{1-\sigma}\frac{1}{\bar{\varphi}_t}\bar{W} \\
&= \left(\frac{1}{r-1}\right)^2 \bar{P}^{-1}\bar{Y}^{1-\sigma}\left[r-(2-r+\varepsilon)\frac{1}{\bar{\varphi}_t^{1+\varepsilon}}\bar{Y}^{\sigma+\varepsilon}\right] \\
&= \left(\frac{1}{r-1}\right)^2 \bar{P}^{-1}\bar{Y}^{1-\sigma} r\left[1-(2-r+\varepsilon)\right] \\
&= \left(\frac{1}{r-1}\right)^2 \bar{P}^{-1}\bar{Y}^{1-\sigma} r\left[r-1-\varepsilon\right]
\end{aligned}
$$

Notice that since $r \in [0,1)$, then this term is strictly negative.

Second derivative, with respect to P_{it}

$$
\begin{aligned}
\frac{\partial^2 \Pi_{jt}}{\partial P_{jt}\partial P_{it}} &= \frac{\partial}{\partial P_{it}} E_{jt}\left(\frac{r}{r-1}\right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma} \\
&\quad - \left(\frac{1}{r-1}\right) P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1+\varepsilon} \frac{1}{\varphi_{it}^{1+\varepsilon}} P_t \left[\left(\frac{P_{jt}}{P_{it}}\right)^{\frac{1}{r-1}} \left(\frac{P_{it}}{P_t}\right)^{\frac{1}{p-1}}\right]^{\varepsilon} \\
&= E_{jt}\left(\frac{r}{r-1}\right)\left(\frac{r-p}{(p-1)(r-1)}\right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}-1} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma} \\
&\quad - \left(\frac{1}{r-1}\right)\left(\frac{(1+\varepsilon)(r-p)}{(p-1)(r-1)}\right) P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}-1} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma} \frac{W_{jt}}{\varphi_{it}}
\end{aligned}
$$

$$
\begin{aligned}
\Pi_{12} \equiv \left.\frac{\partial^2 \Pi_{jt}}{\partial P_{jt}\partial P_{it}}\right|_{\bar{P},\bar{Y},\bar{\varphi}_t} &= \left(\frac{r}{r-1}\right)\left(\frac{r-p}{(p-1)(r-1)}\right)\bar{P}^{-1}\bar{Y}^{1-\sigma} \\
&\quad - \left(\frac{1}{r-1}\right)\left(\frac{(1+\varepsilon)(r-p)}{(p-1)(r-1)}\right)\bar{P}^{-2}\bar{Y}^{1-\sigma}\frac{\bar{W}}{\bar{\varphi}_t} \\
&= \frac{r-p}{(p-1)(r-1)^2}\bar{P}^{-1}\bar{Y}^{1-\sigma}\left[r-(1+\varepsilon)\frac{1}{\bar{\varphi}_t^{1+\varepsilon}}\bar{Y}^{\sigma+\varepsilon}\right] \\
&= \frac{r-p}{(p-1)(r-1)^2}\bar{P}^{-1}\bar{Y}^{1-\sigma} r\left[1-(1+\varepsilon)\right] \\
&= \frac{r-p}{(p-1)(r-1)^2}\bar{P}^{-1}\bar{Y}^{1-\sigma} r\left[-\varepsilon\right]
\end{aligned}
$$

Second derivative, with respect to P_t

$$\frac{\partial^2 \Pi_{jt}}{\partial P_{jt} \partial P_t} = \frac{\partial}{\partial P_t} E_{jt} \left(\frac{r}{r-1} \right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma}$$
$$- \left(\frac{1}{r-1} \right) P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1+\varepsilon} \frac{1}{\varphi_{it}^{1+\varepsilon}} P_t \left[\left(\frac{P_{jt}}{P_{it}} \right)^{\frac{1}{r-1}} \left(\frac{P_{it}}{P_t} \right)^{\frac{1}{p-1}} \right]^{\varepsilon}$$
$$= E_{jt} \left(\frac{r}{r-1} \right) \left(\frac{1}{1-p} \right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}-1} Y_t^{1-\sigma}$$
$$- \left(\frac{1}{r-1} \right) \left(\frac{1}{1-p} + 1 + \frac{\varepsilon}{1-p} \right) P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1+\varepsilon} \frac{W_{jt}}{P_t \varphi_{it}}$$

$$\Pi_{13} \equiv \left. \frac{\partial^2 \Pi_{jt}}{\partial P_{jt} \partial P_t} \right|_{\bar{P}, \bar{Y}, \bar{\varphi}_t} = \left(\frac{r}{r-1} \right) \left(\frac{1}{1-p} \right) \bar{P}^{-1} \bar{Y}^{1-\sigma}$$
$$- \left(\frac{1}{r-1} \right) \left(\frac{2-p+\varepsilon}{1-p} \right) \bar{P}^{-1} \bar{Y}^{1+\varepsilon} \frac{\bar{W}}{\bar{P} \bar{\varphi}_t}$$
$$= \frac{1}{(1-p)(r-1)} \bar{P}^{-1} \bar{Y}^{1-\sigma} \left[r - (2-p+\varepsilon) \frac{1}{\bar{\varphi}_t^{1+\varepsilon}} \bar{Y}^{\sigma+\varepsilon} \right]$$
$$= -\frac{1}{(p-1)(r-1)} \bar{P}^{-1} \bar{Y}^{1-\sigma} r \left[-1 + p - \varepsilon \right]$$
$$= (-1) \alpha^{-1} \frac{1}{r-1} \bar{P}^{-1} \bar{Y}^{1-\sigma} r$$

Second derivative, with respect to Y_t

$$\frac{\partial^2 \Pi_{jt}}{\partial P_{jt} \partial Y_t} = \frac{\partial}{\partial Y_t} E_{jt} \left(\frac{r}{r-1} \right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma}$$
$$- \left(\frac{1}{r-1} \right) P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1+\varepsilon} \frac{1}{\varphi_{it}^{1+\varepsilon}} P_t \left[\left(\frac{P_{jt}}{P_{it}} \right)^{\frac{1}{r-1}} \left(\frac{P_{it}}{P_t} \right)^{\frac{1}{p-1}} \right]^{\varepsilon}$$
$$= E_{jt} \left(\frac{r}{r-1} \right) (1-\sigma) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{-\sigma}$$
$$- \left(\frac{1}{r-1} \right) (1+\varepsilon) P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{\varepsilon} \frac{1}{\varphi_{it}^{1+\varepsilon}} P_t$$
$$\times \left[\left(\frac{P_{jt}}{P_{it}} \right)^{\frac{1}{r-1}} \left(\frac{P_{it}}{P_t} \right)^{\frac{1}{p-1}} \right]^{\varepsilon}$$

$$\Pi_{14} \equiv \left.\frac{\partial^2 \Pi_{jt}}{\partial P_{jt} \partial Y_t}\right|_{\bar{P},\bar{Y},\bar{\varphi}_t} = \left(\frac{r}{r-1}\right)(1-\sigma)\bar{Y}^{-\sigma}$$

$$- \left(\frac{1}{r-1}\right)(1+\varepsilon)\bar{P}^{-1}\bar{Y}^{\varepsilon}\frac{1}{\bar{\varphi}_t^{1+\varepsilon}}$$

$$= \frac{1}{r-1}\bar{Y}^{-\sigma}\left[(1-\sigma)r - (1+\varepsilon)\frac{1}{\bar{\varphi}_t^{1+\varepsilon}}\bar{Y}^{\sigma+\varepsilon}\right]$$

$$= \frac{1}{r-1}\bar{Y}^{-\sigma} r\left[(1-\sigma) - (1+\varepsilon)\right]$$

$$= (-1)\frac{1}{r-1}\bar{Y}^{-\sigma} r\left[\varepsilon + \sigma\right]$$

Second derivative, with respect to φ_t

$$\frac{\partial^2 \Pi_{jt}}{\partial P_{jt} \partial \varphi_t} = \frac{\partial}{\partial \varphi_{it}} E_{jt}\left(\frac{r}{r-1}\right) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1-\sigma}$$

$$- \left(\frac{1}{r-1}\right) P_{jt}^{\frac{1}{r-1}-1} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1+\varepsilon} \frac{1}{\varphi_{it}^{1+\varepsilon}} P_t \left[\left(\frac{P_{jt}}{P_{it}}\right)^{\frac{1}{r-1}} \left(\frac{P_{it}}{P_t}\right)^{\frac{1}{p-1}}\right]^{\varepsilon}$$

$$= \left(\frac{1}{r-1}\right)(1+\varepsilon) P_{jt}^{\frac{1}{r-1}} P_{it}^{\frac{r-p}{(p-1)(r-1)}} P_t^{\frac{1}{1-p}} Y_t^{1+\varepsilon} \frac{1}{\varphi^{2+\varepsilon}}$$

$$\times \left[\left(\frac{P_{jt}}{P_{it}}\right)^{\frac{1}{r-1}} \left(\frac{P_{it}}{P_t}\right)^{\frac{1}{p-1}}\right]^{\varepsilon}$$

$$\Pi_{15} \equiv \left.\frac{\partial^2 \Pi_{jt}}{\partial P_{jt} \partial \varphi_t}\right|_{\bar{P},\bar{Y},\bar{\varphi}_t} = \left(\frac{1}{r-1}\right)(1+\varepsilon)\bar{Y}^{1+\varepsilon+\sigma-\sigma}\frac{1}{\bar{\varphi}_t^{2+\varepsilon}}$$

$$= \left(\frac{1+\varepsilon}{r-1}\right) r\bar{Y}^{1-\sigma}\frac{1}{\bar{\varphi}_t}$$

Log-linear approximation to an intermediate good firm's pricing decision

Taking first-order conditions with respect to the firm's objective function (A.3) yields

$$
\begin{aligned}
\frac{\partial \tilde{\Pi}}{\partial \tilde{p}_{jt}} = 0 =& \Pi_1 \bar{P} + \Pi_{11} \bar{P}^2 \tilde{p}_{jt} + \Pi_{12} \bar{P}^2 \tilde{p}_{it} + \Pi_{13} \bar{P}^2 E_{jt} \tilde{p}_t \\
& + \Pi_{14} \bar{P} \bar{Y} E_{jt} \tilde{y}_t + \Pi_{15} \bar{P} \bar{\varphi}_t \tilde{E}_{jt} \tilde{\phi}_{it} \\
=& 0 + \Pi_{11} \bar{P}^2 \tilde{p}_{jt} + \Pi_{12} \bar{P}^2 \tilde{p}_{it} \\
& + \Pi_{13} \bar{P}^2 E_{jt} \tilde{p}_t + \Pi_{14} \bar{P} \bar{Y} E_{jt} \tilde{y}_t \\
& + \Pi_{15} \bar{P} \bar{\varphi}_t E_{jt} \tilde{\phi}_{it} \\
=& 0 \\
& + \left(\frac{1}{r-1} \right)^2 \bar{P}^{-1} \bar{Y}^{1-\sigma} r \left[r - 1 - \varepsilon \right] \bar{P}^2 \tilde{p}_{jt} \\
& + \frac{r-p}{(p-1)(r-1)^2} \bar{P}^{-1} \bar{Y}^{1-\sigma} r \left[-\varepsilon \right] \bar{P}^2 \tilde{p}_{it} \\
& + (-1) \alpha^{-1} \frac{1}{r-1} \bar{P}^{-1} \bar{Y}^{1-\sigma} r \bar{P}^2 E_{jt} \tilde{p}_t \\
& + (-1) \frac{1}{r-1} \bar{Y}^{-\sigma} r \left[\varepsilon + \sigma \right] \bar{P} \bar{Y} E_{jt} \tilde{y}_t \\
& + \left(\frac{1+\varepsilon}{r-1} \right) r \bar{Y}^{1-\sigma} \frac{1}{\bar{\varphi}_t} \bar{P} \bar{\varphi}_t E_{jt} \tilde{\phi}_{it} \\
=& + \left(\frac{1}{r-1} \right)^2 \left[r - 1 - \varepsilon \right] \tilde{p}_{jt} \\
& + \frac{r-p}{(p-1)(r-1)^2} \left[-\varepsilon \right] \tilde{p}_{it} \\
& + (-1) \alpha^{-1} \frac{1}{r-1} E_{jt} \tilde{p}_t \\
& + (-1) \frac{1}{r-1} \left[\varepsilon + \sigma \right] E_{jt} \tilde{y}_t \\
& + \left(\frac{1+\varepsilon}{r-1} \right) E_{jt} \tilde{\phi}_{it}
\end{aligned}
$$

Notice that $P_{it} = P_{jt}$ since all firms within an industry face the same problem, and also that:

$$\begin{aligned}(p-1)(r-1-\varepsilon)+(r-p)(-\varepsilon) &= pr-p-p\varepsilon-r+1+\varepsilon-r\varepsilon+p\varepsilon \\ &= r(p-1-\varepsilon)-(p-1-\varepsilon) \\ &= (r-1)(p-1-\varepsilon)\end{aligned}$$

and $\frac{p-1-\varepsilon}{p-1} = \frac{1-p+\varepsilon}{1-p} \equiv \alpha^{-1}$. Then we have

$$\begin{aligned}0 =& \frac{1}{(p-1)(r-1)^2}\left[(p-1)(r-1-\varepsilon)+(r-p)(-\varepsilon)\right]\tilde{p}_{jt} \\ &+(-1)\alpha^{-1}\frac{1}{r-1}E_{jt}\tilde{p}_t \\ &+(-1)\frac{1}{r-1}\left[\varepsilon+\sigma\right]E_{jt}\tilde{y}_t \\ &+\left(\frac{1+\varepsilon}{r-1}\right)E_{jt}\tilde{\phi}_{it} \\ =&\alpha^{-1}\tilde{p}_{jt} \\ &+(-1)\alpha^{-1}E_{jt}\tilde{p}_t \\ &+(-1)(\varepsilon+\sigma)E_{jt}\tilde{y}_t \\ &+(1+\varepsilon)E_{jt}\tilde{\phi}_{it}\end{aligned}$$

and finally this reduces to the firms' imperfect-information pricing rule

$$\begin{aligned}\tilde{p}^*_{jt} &= -\gamma E_{jt}\tilde{\phi}_{it}+\zeta E_{jt}\tilde{y}_t+E_{jt}\tilde{p}_t \\ &= -\gamma E_{jt}\tilde{\phi}_{it}+\zeta E_{jt}\tilde{q}_t+(1-\zeta)E_{jt}\tilde{p}_t \\ &= E_{jt}p^{\diamond}_{jt}\end{aligned}$$

Log-quadatic approximation to profit loss due to imperfect information

We cconsider the loss in profit from a firm setting a non-profit-maximizing price $p^*_{jt} = E_{jt} p^\diamond_{jt}$. The log-quadratic approximation to the profit function is

$$\begin{aligned}\tilde{\Pi}(\tilde{p}_{jt}, \tilde{p}_{it}, \tilde{p}_t, \tilde{y}_t, \tilde{\phi}_{it}) =& \Pi_1 \bar{P} \tilde{p}_{jt} + \frac{\Pi_{11}}{2!} \bar{P}^2 \tilde{p}^2_{jt} + \Pi_{12} \bar{P}^2 \tilde{p}_{jt} \tilde{p}_{it} + \Pi_{13} \bar{P}^2 \tilde{p}_{jt} E_{jt} \tilde{p}_t \\ & + \Pi_{14} \bar{P} \bar{Y} \tilde{p}_{jt} E_{jt} \tilde{y}_t + \Pi_{15} \bar{P} \bar{\varphi}_t \tilde{p}_{jt} E_{jt} \tilde{\phi}_{it} \\ & + \text{other terms}\end{aligned}$$

and recall that the "other terms" do not depend on the firm's price decision. The loss in profits is then

$$\begin{aligned}\tilde{\Pi}(\tilde{p}^\diamond_{jt}, \cdot) - \tilde{\Pi}(\tilde{p}^*_{jt}, \cdot) =& \Pi_1 \bar{P} (\tilde{p}^\diamond_{jt} - \tilde{p}^*_{jt}) \\ & + \left(\frac{\Pi_{11}}{2} \right) \bar{P}^2 \left(\tilde{p}^{\diamond^2}_{jt} - \tilde{p}^{*^2}_{jt} \right) \\ & + (\Pi_{12} \bar{P}^2 \tilde{p}_{it} + \Pi_{13} \bar{P}^2 \tilde{p}_t + \Pi_{14} \bar{P} \bar{Y} \tilde{Y}_t + \Pi_{15} \bar{P} \bar{\varphi}_t \tilde{\phi}_{it}) \\ \times & (\tilde{p}^\diamond_{jt} - \tilde{p}^*_{jt})\end{aligned}$$

Note first that $\Pi_1 = 0$, and second, from the first-order condition above in the perfect-information case, that $\Pi_{12} \bar{P}^2 \tilde{p}_{it} + \Pi_{13} \bar{P}^2 \tilde{p}_t + \Pi_{14} \bar{P} \bar{Y} \tilde{Y}_t + \Pi_{15} \bar{P} \bar{\varphi}_t \tilde{\phi}_t = -\Pi_{11} \bar{P}^2 \tilde{p}^\diamond_{jt}$. Then we can rewrite the loss in profits as

$$\begin{aligned}\tilde{\Pi}(\tilde{p}^\diamond_{jt}, \cdot) - \tilde{\Pi}(\tilde{p}^*_{jt}, \cdot) &= \bar{P}^2 \left(\frac{\Pi_{11}}{2} \right) \left(\tilde{p}^{\diamond^2}_{jt} - \tilde{p}^{*^2}_{jt} \right) - \bar{P}^2 \Pi_{11} \tilde{p}^\diamond_{jt} (\tilde{p}^\diamond_{jt} - \tilde{p}^*_{jt}) \\ &= -\bar{P}^2 \left(\frac{\Pi_{11}}{2} \right) \left(\tilde{p}^{\diamond^2}_{jt} + \tilde{p}^{*^2}_{jt} \right) + \bar{P}^2 \Pi_{11} \tilde{p}^\diamond_{jt} \tilde{p}^*_{jt} \\ &= \left(-\frac{\Pi_{11}}{2} \bar{P}^2 \right) (\tilde{p}^\diamond_{jt} - \tilde{p}^*_{jt})^2\end{aligned}$$

Finally recall, from above, that $\Pi_{11} < 0$ so that the above is positive overall, indicating setting $\tilde{p}^{*}_{jt}$ yields less profits than setting $\tilde{p}^{\diamond}_{jt}$. The expected loss in profits due to imperfect information can be written

$$E_{jt}\left[\tilde{\Pi}(\tilde{p}^{\diamond}_{jt},\cdot)-\tilde{\Pi}(\tilde{p}^{*}_{jt},\cdot)\right]=-\hat{\Pi}_{11}\left(\tilde{p}^{\diamond}_{jt}-\tilde{p}^{*}_{jt}\right)^{2}$$

where $\hat{\Pi}_{11}=\left(\frac{\Pi_{11}}{2}\bar{P}^{2}\right)<0$.

Information Theory

This appendix collects information theoretic results.

Mutual Information of Random Vectors

In the case that the variables are independent so that is no reduction in uncertainty, then $I(X;S) = H(X) - H(X) = 0$. Supposing that $\mathbf{X}$ and $\mathbf{S}$ are finite n-dimensional independent vectors such that X_i and S_j are independent if $i \neq j$, then

$$\begin{aligned}
I(\mathbf{X};\mathbf{S}) &= H(X_1,\cdots,X_n) - H(X_1,\cdots,X_n|S_1,\cdots,S_n)\\
&= \sum_{i=1}^{n} H(X_i) + \sum_{i=1}^{n} H(X_i) - H(X_1,\cdots,X_n,S_1,\cdots,S_n)\\
&= \sum_{i=1}^{n} H(X_i) + \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i,S_i)\\
&= \sum_{i=1}^{n} I(X_i;S_i)
\end{aligned}$$

where the third equality follows from an iterative application of the chain rule.

Gaussian Mutual Information

Suppose, as we do above, that we have two mutually Gaussian random variables, ω and s^{ω} such that

$$s^{(\omega)} = \omega + \psi$$

where ψ is Gaussian white noise. A well-known result (see for example Cover and Thomas, 2006) for Gaussian random processes is that mutual information can be simply expressed.

$$\mathcal{I}(\omega, s^{(\omega)}) = \frac{1}{2}\log_2\left(\frac{1}{1-\rho^2_{\omega s^{(\omega)}}}\right)$$

where $\rho^2_{\omega s^{(\omega)}}$ is the correlation coefficent between the two processes. Now, notice that the correlation coefficient can be rewritten in terms of the processes variances

$$\begin{aligned}\rho^2_{\omega s^{(\omega)}} &= \left[\frac{Cov(\omega, s^{(\omega)})}{\sigma_\omega \sigma_{s^{(\omega)}}}\right]^2 = \left[\frac{\sigma^2_\omega}{\sigma_\omega \sigma_{s^{(\omega)}}}\right]^2 = \frac{\sigma^2_\omega}{\sigma^2_{s^{(\omega)}}} \\ &= \frac{\sigma^2_\omega}{\sigma^2_\omega + \sigma^2_\psi}\end{aligned}$$

and $1 - \rho^2_{\omega s^{(\omega)}} = \frac{\sigma_\psi}{\sigma^2_\omega + \sigma^2_\psi}$. Then the mutual information can be rewritten also in terms of the processes variances

$$\begin{aligned}\mathcal{I}(\omega, s^{(\omega)}) &= \frac{1}{2}\log_2\left(\frac{\sigma^2_\omega + \sigma^2_\psi}{\sigma^2_\psi}\right) \\ &= \frac{1}{2}\log_2\left(\frac{\sigma^2_\omega}{\sigma^2_\psi} + 1\right)\end{aligned}$$

Expressions for Mutual Information

Define the mutual information as $\kappa_j^{(\omega)} \equiv \mathcal{I}(\omega, s_j^{(\omega)})$. Assume the processes are defined as above.

1. Using this, we can find an expression for the ratio of the variances in terms as a function of a level of mutual information:

$$2^{2\kappa_j^{(\omega)}} - 1 = \frac{\sigma_\omega^2}{\sigma_\psi^2}$$

2. Immediately we also have an expression for the variance of the signal in terms of a level of mutual information and the variance of the fundamental

$$\sigma_\psi^2 = \left(2^{2\kappa_j^{(\omega)}} - 1\right)^{-1} \sigma_\omega^2$$

3. The ratio of the variance of the fundamental to the variance of the signal, a key term in typical signal extraction results, can be derived from (2)

$$\begin{aligned} 2^{2\kappa_j^{(\omega)}} \sigma_\psi^2 &= \sigma_\omega^2 + \sigma_\psi^2 \\ 2^{-2\kappa_j^{(\omega)}} &= \frac{\sigma_\psi^2}{\sigma_\omega^2 + \sigma_\psi^2} \\ 1 - 2^{-2\kappa_j^{(\omega)}} &= \frac{\sigma_\omega^2}{\sigma_\omega^2 + \sigma_\psi^2} \end{aligned}$$

4. Finally, from (1) - (3) follows a result that will useful in expanding the expected loss of profits from setting a price with imperfect information

$$\left(1-2^{-2\kappa_j^{(\omega)}}\right)^2 = \left(\frac{2^{2\kappa_j^{(\omega)}}-1}{2^{2\kappa_j^{(\omega)}}}\right)^2$$
$$\left(1-2^{-2\kappa_j^{(\omega)}}\right)^2 \sigma_\psi^2 = \left(\frac{2^{2\kappa_j^{(\omega)}}-1}{2^{2\kappa_j^{(\omega)}}}\right)^2 \left(2^{2\kappa_j^{(\omega)}}-1\right)^{-1} \sigma_\omega^2$$
$$\left(1-2^{-2\kappa_j^{(\omega)}}\right)^2 \sigma_\psi^2 = \left(\frac{2^{2\kappa_j^{(\omega)}}-1}{2^{4\kappa_j^{(\omega)}}}\right) \sigma_\omega^2$$

Equilibrium

Perfect Information

In the perfect information case, firms set prices according to (see *Optimal Price Setting*):

$$p_{jt}^{\diamond} = \alpha \log \frac{1}{r} - \gamma\phi_{it} + \zeta q_t + (1-\zeta)p_t$$

Applying symmetry across firms within each industry, integrating across all firms, and using the log-linear approximation to the aggregate price index around the point where all prices are the same (see *Log-linear approximation to the aggregate* *price index*) yields

$$\begin{aligned}
\int_J p_{jt} dj &= \int_J \alpha \log \frac{1}{r} dj - \int_J \gamma \phi_{it} dj + \int_J \zeta y_t dj + \int_J p_t dj \\
\sum_{l=1}^{I} \mu_l p_{lt} &= \alpha \log \frac{1}{r} - \gamma \sum_{l=1}^{I} \mu_l \phi_{lt} + \zeta y_t + p_t \\
p_t &= \alpha \log \frac{1}{r} - \gamma \sum_{l=1}^{I} \mu_l \phi_{lt} + \zeta y_t + p_t
\end{aligned}$$

The perfect information equilibrium level of (real) output is then

$$y_t^{\diamond} = -\frac{\alpha}{\zeta} \log \frac{1}{r} + \frac{\gamma}{\zeta} \sum_{l=1}^{I} \mu_l \phi_{lt}$$

If nominal output is assumed to be an exogenous process $Q_t = P_t Y_t$ then the perfect information equilibrium aggregate price level is

$$\begin{aligned}
p_t^{\diamond} &= q_t - y_t^n \\
&= q_t + \frac{\alpha}{\zeta} \log \frac{1}{r} - \frac{\gamma}{\zeta} \sum_{l=1}^{I} \mu_l \phi_{lt}
\end{aligned}$$

In deviation from steady-state form, these are written

$$\tilde{y}_t^{\diamond} = \frac{\gamma}{\zeta} \sum_{l=1}^{I} \mu_l \tilde{\phi}_{lt}$$

$$\tilde{p}_t^{\diamond} = \tilde{q}_t - \frac{\gamma}{\zeta} \sum_{l=1}^{I} \mu_l \tilde{\phi}_{lt}$$

And plugging this into the firms' pricing decision yields

$$\tilde{p}^{\diamond}_{jt} = -\gamma\tilde{\phi}_{it} + \zeta\tilde{q}_t + (1-\zeta)p^{\diamond}_t$$

<u>Non-stochastic equilibrium</u>

In the non-stochastic case where shocks are set to their means, we have

$$y^n_t = -\frac{\alpha}{\zeta}\log\frac{1}{r} + \frac{\gamma}{\zeta}\sum_{l=1}^{I}\mu_l\bar{\phi}_{lt}$$

$$p^n_t = \bar{q}_t + \frac{\alpha}{\zeta}\log\frac{1}{r} - \frac{\gamma}{\zeta}\sum_{l=1}^{I}\mu_l\bar{\phi}_{lt}$$

Plug in the aggregate price level to the firm's pricing rule

$$\begin{aligned}
p^n_{jt} &= \alpha\log\frac{1}{r} - \gamma\bar{\phi}_{it} + \zeta\bar{q}_t + (1-\zeta)\left[\bar{q}_t + \frac{\alpha}{\zeta}\log\frac{1}{r} - \frac{\gamma}{\zeta}\sum_{l=1}^{I}\mu_l\bar{\phi}_{lt}\right] \\
&= \left(1 + \frac{1-\zeta}{\zeta}\right)\alpha\log\frac{1}{r} - \gamma\bar{\phi}_{it} + (\zeta + 1 - \zeta)\bar{q}_t + (1-\zeta)\frac{\gamma}{\zeta}\sum_{l=1}^{I}\mu_l\bar{\phi}_{lt} \\
&= \bar{q}_t + \frac{\alpha}{\zeta}\log\frac{1}{r} - (1-\zeta)\frac{\gamma}{\zeta}\sum_{l=1}^{I}\mu_l\bar{\phi}_{lt} - \gamma\bar{\phi}_{it}
\end{aligned}$$

which yields the non-stochastic equilibrium price rule

$$p^n_{jt} = p^n_t + \gamma\left[\sum_{l=1}^{I}\mu_l\bar{\phi}_{lt} - \bar{\phi}_{it}\right]$$

To confirm that this is an equilibrium, integrate this pricing rule over all firms and use the log-linear approximation to the aggregate price index.

Notice that if all shocks are eliminated so that $\phi_{it} = 0$ for $i = 1, \ldots, I$ or if the shock is purely aggregate so that $\phi_{it} = \phi_{i't}$ for $i, i' = 1, \ldots, I$ then the perfect information equilibrium corresponds to point where all prices are the same and are equal to the aggregate price index (this last point is by construction, see *Budget Contraints*).

Rational Inattention under Gaussian White Noise

Here we follow a guess and verify approach. Given the form of the perfect information equilibrium, we guess that the equilibrium aggregate price level is given by

$$\tilde{p}_t^* = a\tilde{q}_t - \frac{\gamma}{\zeta} \sum_{l=1}^{I} b_l \mu_l \tilde{\phi}_{lt}$$

where a and $\{b_l\}_{l=1}^{I}$ are coefficients governing the extent of adjustment of the price level due to shocks.

<u>Imperfect Information Pricing Rule</u>

With this guess, firms' imperfect-information optimal price rule is

$$\begin{aligned}
\tilde{p}_{jt}^* &= E_{jt}\tilde{p}_{jt}^{\diamond} \\
&= -\gamma E_{jt}\tilde{\phi}_{it} + \zeta E_{jt} q_t + (1-\zeta) E_{jt}\tilde{p}_t \\
&= -\gamma E_{jt}\tilde{\phi}_{it} + \zeta E_{jt} q_t + (1-\zeta)\left[a E_{jt}\tilde{q}_t - \frac{\gamma}{\zeta} \sum_{l=1}^{I} b_l \mu_l E_{jt}\tilde{\phi}_{lt} \right] \\
&= \left[(1-\zeta)a + \zeta\right] E_{jt}\tilde{q}_t - (1-\zeta)\frac{\gamma}{\zeta} \sum_{l=1}^{I} b_l \mu_l \tilde{\phi}_{lt} - \gamma E_{jt}\tilde{\phi}_{it}
\end{aligned}$$

To ease notation in the following optimization problem in which the constant term simply be carried around, define $\bar{a} \equiv [(1-\zeta)a + \zeta]$ and $\bar{b}_l \equiv (1-\zeta)\frac{\gamma}{\zeta}b_l$; they will be unpacked again at the end to aid interpretation. Then the optimal price rule is

$$\tilde{p}^*_{jt} = \bar{a}E_{jt}\tilde{q}_t - \sum_{l=1}^{I} \bar{b}_l \mu_l \tilde{\phi}_{lt} - \gamma E_{jt}\tilde{\phi}_{it}$$

Price-Setting Mean Squared Error

Given the signals the firm receives, we can solve the expectations using typical signal extraction results

$$\tilde{p}^*_{jt} = \bar{a}\left(\frac{\sigma_q^2}{\sigma_q^2 + \sigma^2_{\psi_j^{(q)}}}\right) s^q_{jt} - \sum_{l \neq i} \bar{b}_l \mu_l \left(\frac{\sigma_l^2}{\sigma_l^2 + \sigma^2_{\psi_j^{(l)}}}\right) s^{(l)}_{jt}$$
$$- (\bar{b}_i \mu_i + \gamma)\left(\frac{\sigma_i^2}{\sigma_i^2 + \sigma^2_{\psi_j^{(i)}}}\right) s^{(i)}_{jt}$$

Using results from *Expressions for Mutual Information*, we can rewrite this in terms of mutual information as

$$\begin{aligned}\tilde{p}^*_{jt} =& \bar{a}\left(1 - 2^{-2\kappa_j^{(q)}}\right)\left(\tilde{q}_t - \psi^{(q)}_{jt}\right) \\ &- \sum_{l \neq i} \bar{b}_l \mu_l \left(1 - 2^{-2\kappa_j^{(l)}}\right)\left(\tilde{\phi}_{lt} - \psi^{(l)}_{jt}\right) \\ &- (\bar{b}_i \mu_i + \gamma)\left(1 - 2^{-2\kappa_j^{(i)}}\right)\left(\tilde{\phi}_{it} - \psi^{(i)}_{jt}\right)\end{aligned}$$

The expected loss in profits from setting an imperfect-information price is

$$E_{jt}\left[\tilde{\Pi}_{jt}\left(\tilde{p}^{\diamond}_{jt},\cdot\right)-\tilde{\Pi}_{jt}\left(\tilde{p}^{*}_{jt},\cdot\right)\right]=\left(\frac{\hat{\Pi}_{11}}{2}\right)E_{jt}\left[\left(\tilde{p}^{\diamond}_{jt}-\tilde{p}^{*}_{jt}\right)^{2}\right]$$

which is a constant times the mean squared error of the imperfect-information price. The difference between the perfect- and imperfect-information prices is

$$\begin{aligned}\tilde{p}^{\diamond}_{jt}-\tilde{p}^{*}_{jt}=&\bar{a}\left[\tilde{q}_t-\left(1-2^{-2\kappa_j^{(q)}}\right)\left(\tilde{q}_t+\psi^{(q)}_{jt}\right)\right]\\&-\sum_{l\neq i}\bar{b}_l\mu_l\left[\tilde{\phi}_{lt}-\left(1-2^{-2\kappa_j^{(l)}}\right)\left(\tilde{\phi}_{lt}+\psi^{(l)}_{jt}\right)\right]\\&-\left(\bar{b}_i\mu_i+\gamma\right)\left[\tilde{\phi}_{it}-\left(1-2^{-2\kappa_j^{(i)}}\right)\left(\tilde{\phi}_{it}+\psi^{(i)}_{jt}\right)\right]\\=&\bar{a}\left[2^{-2\kappa_j^{(q)}}\tilde{q}_t-\left(1-2^{-2\kappa_j^{(q)}}\right)\psi^{(q)}_{jt}\right]\\&-\sum_{l\neq i}\bar{b}_l\mu_l\left[2^{-2\kappa_j^{(l)}}\tilde{\phi}_{lt}-\left(1-2^{-2\kappa_j^{(l)}}\right)\psi^{(l)}_{jt}\right]\\&-\left(\bar{b}_i\mu_i+\gamma\right)\left[2^{-2\kappa_j^{(i)}}\tilde{\phi}_{it}-\left(1-2^{-2\kappa_j^{(i)}}\right)\psi^{(i)}_{jt}\right]\end{aligned}$$

then noting that independence implies that all cross terms have expected value zero, the mean squared error can be expressed

$$\begin{aligned}E_{jt}\left[\left(\tilde{p}^{\diamond}_{jt}-\tilde{p}^{*}_{jt}\right)^{2}\right]=&\bar{a}^2\left[2^{-4\kappa_j^{(q)}}\sigma^2_q+\left(1-2^{-2\kappa_j^{(q)}}\right)^2\sigma^2_{\psi_j^{(q)}}\right]\\&+\sum_{l=1}^{I}\bar{b}_l^2\mu_l^2\left[2^{-4\kappa_j^{(l)}}\sigma^2_{\phi_l}+\left(1-2^{-2\kappa_j^{(l)}}\right)^2\sigma^2_{\psi_j^{(l)}}\right]\\&+\left(\bar{b}_i\mu_i+\gamma\right)^2\left[2^{-4\kappa_j^{(i)}}\sigma^2_{\phi_i}+\left(1-2^{-2\kappa_j^{(i)}}\right)^2\sigma^2_{\psi_j^{(i)}}\right]\end{aligned}$$

then using result (4) from *Expressions for Mutual Information*, it can be finally written

$$E_{jt}\left[\left(\tilde{p}^{\diamond}_{jt}-\tilde{p}^{*}_{jt}\right)^2\right]=\bar{a}^2 2^{-2\kappa_j^{(q)}}\sigma_q^2+\sum_{l=1}^{I}\bar{b}_l^2\mu_l^2 2^{-2\kappa_j^{(l)}}\sigma_{\phi_l}^2+(\bar{b}_i\mu_i+\gamma)^2 2^{-2\kappa_j^{(i)}}\sigma_{\phi_i}^2$$

The Attention Problem

The firm's attention problem can now be fully specified

$$\min_{\{\kappa_j^{(\omega)}\}_{\omega\in\Omega}}\sum_{\omega\in\Omega}\left(\bar{\kappa}_j^{(\omega)}\right)^2 2^{-2\kappa_j^{(\omega)}}\qquad;\qquad \bar{\kappa}_j^{(\omega)}=\begin{cases}\bar{a}\sigma_q & \omega=q\\ \bar{b}_l\mu_l\sigma_{\phi_l} & \omega=l\neq i\\ (\bar{b}_i\mu_i+\gamma)\sigma_{\phi_i} & \omega=i\end{cases}$$

such that $\sum_{\omega\in\Omega}\kappa_j^{(\omega)}\leq\kappa$ and $\kappa_j^{(\omega)}\geq 0$. This is a constrained optimization problem and can be represented as a Lagrangian (where the constraint is assumed to be binding, since in any optimum firms will use all available attention)

$$\mathcal{L}=\sum_{\omega\in\Omega}\left(\bar{\kappa}_j^{(\omega)}\right)^2 2^{-2\kappa_j^{(\omega)}}-\lambda\left[\sum_{\omega\in\Omega}\kappa_j^{(\omega)}-\kappa\right]$$

Assuming an interior solution, the $|\Omega|$ first-order conditions for an optimum are

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\kappa_j^{(\omega)}}=0&=\left(\bar{\kappa}_j^{(\omega)}\right)^2 2^{-2\kappa_j^{(\omega)}}(-2\ln 2)-\lambda\\ 2^{2\kappa_j^{(\omega)}}&=\frac{\left(\bar{\kappa}_j^{(\omega)}\right)^2(-2\ln 2)}{\lambda}\\ \kappa_j^{(\omega)}&=\frac{1}{2}\log_2\left(\frac{-2\ln 2}{\lambda}\right)+\frac{1}{2}\log_2\left[\left(\bar{\kappa}_j^{(\omega)}\right)^2\right]\end{aligned}$$

Define $\hat{\kappa}_k^{(\omega)} = \frac{1}{2}\log_2\left[\left(\bar{\kappa}_j^{(\omega)}\right)^2\right]$ to ease notation and use the first condition to substitute out the Lagrange multiplier

$$\frac{1}{2}\log_2\left(\frac{-2\ln 2}{\lambda}\right) = \kappa_j^{(\omega_1)} - \hat{\kappa}_j^{(\omega_1)}$$

Then use this in the remaining $|\Omega|-1$ conditions to get:

$$\begin{aligned} \kappa_j^{(\omega_k)} &= \kappa_j^{(\omega_1)} - \hat{\kappa}_j^{(\omega_1)} + \hat{\kappa}_j^{(\omega_k)} \\ -\kappa_j^{(\omega_1)} + \kappa_j^{(\omega_k)} &= -\hat{\kappa}_j^{(\omega_1)} + \hat{\kappa}_j^{(\omega_k)} \qquad\qquad k = 2,\ldots,|\Omega| \end{aligned}$$

Including the constraint $\sum_{\omega\in\Omega}\kappa_j^{(\omega)} = \kappa$ there are $|\Omega|$ equations and $|\Omega|$ unknowns. This can be written in the following linear system:

$$\begin{bmatrix} -1 & 1 & 0 & \cdots & & 0 \\ -1 & 0 & 1 & \cdots & & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ -1 & \cdots & & 0 & 1 & 0 \\ -1 & \cdots & & & 0 & 1 \\ 1 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \kappa_j^{(\omega_1)} \\ \kappa_j^{(\omega_2)} \\ \vdots \\ \kappa_j^{(\omega_{|\Omega|-1})} \\ \kappa_j^{(\omega_{|\Omega|})} \end{bmatrix} = \begin{bmatrix} -\hat{\kappa}_j^{(\omega_1)} + \hat{\kappa}_j^{(\omega_2)} \\ -\hat{\kappa}_j^{(\omega_1)} + \hat{\kappa}_j^{(\omega_3)} \\ \vdots \\ -\hat{\kappa}_j^{(\omega_1)} + \hat{\kappa}_j^{(\omega_{|\Omega|})} \\ \kappa \end{bmatrix}$$

This can be solved using the following steps:

1. Multiply row 1 by -1
2. For rows 2 through $|\Omega|$, iteratively add the previous row and multiply by -1
3. For rows $l = 1,\ldots,|\Omega|-1$, substract l times the l^{th} for from row $|\Omega|$.
4. Divide row $|\Omega|$ by $|\Omega|$

5. For rows $l = |\Omega| - 1, \ldots, 1$, add the $l + 1^{\text{th}}$ row

6. Simplify

This process yields optimal interior allocations that can be expressed as

$$\kappa_j^{(\omega)^*} = |\Omega|^{-1} \left[\kappa - \sum_{\omega' \neq \omega} \hat{\kappa}_j^{(\omega')} + (|\Omega| - 1)\hat{\kappa}_j^{(\omega)} \right] \qquad \omega \in \Omega$$

We can abuse notation to take into account the corner conditions

$$\kappa_j^{(\omega)^*} = \begin{cases} \kappa & \kappa_j^{(\omega)^*} > \kappa \\ \kappa_j^{(\omega)^*} & \kappa_j^{(\omega)^*} \in [0, \kappa] \\ 0 & \kappa_j^{(\omega)^*} < 0 \end{cases}$$

The allocations can be rewritten

$$\begin{aligned} \kappa_j^{(\omega)^*} &= |\Omega|^{-1} \left[\kappa - \sum_{\omega' \in \Omega} \hat{\kappa}_j^{(\omega')} + |\Omega|\hat{\kappa}_j^{(\omega)} \right] \\ &= |\Omega|^{-1} \left[\frac{1}{2}\log_2\left(2^{2\kappa}\right) - \sum_{\omega' \in \Omega} \frac{1}{2}\log_2\left[\left(\bar{\kappa}_j^{(\omega')} \right)^2 \right] + |\Omega| \frac{1}{2}\log_2\left[\left(\kappa_j^{(\omega)} \right)^2 \right] \right] \end{aligned}$$

defining $\bar{\kappa} = \frac{\kappa}{|\Omega|}$ and $\bar{\kappa}_j = \left[\prod_{\omega' \in \Omega} \bar{\kappa}_j^{(\omega')} \right]^{\frac{1}{|\Omega|}}$, we have the final expression for the optimal allocation of attention

$$\kappa_j^{(\omega)^*} = \log_2 2^{\bar{\kappa}} + \log_2 \bar{\kappa}_j^{(\omega)} - \log_2 \bar{\kappa}_j \qquad \omega \in \Omega$$

Note that it is straightforward that

$$1 - 2^{-2\kappa_j^{(\omega)^*}} = 1 - \left(2^{-2\bar{\kappa}}\right) \left(\bar{\kappa}_j^{(\omega)} \right)^{-2} \bar{\kappa}_j^2$$

Verifying the Guess

Returning to the optimal imperfect-information pricing rule

$$
\begin{aligned}
\tilde{p}_{jt}^{*} = & \bar{a}\left(1-2^{-2\kappa_j^{(q)*}}\right)\left(\tilde{q}_t-\psi_{jt}^{(q)}\right) \\
& -\sum_{l\neq i}\bar{b}_l\mu_l\left(1-2^{-2\kappa_j^{(l)*}}\right)\left(\tilde{\phi}_{lt}-\psi_{jt}^{(l)}\right) \\
& -\left(\bar{b}_i\mu_i+\gamma\right)\left(1-2^{-2\kappa_j^{(i)*}}\right)\left(\tilde{\phi}_{it}-\psi_{jt}^{(i)}\right)
\end{aligned}
$$

Integrating over all firms, applying symmetry to within-industry firms' optimal attention allocations, and applying the log-linear approximation to the aggregate price yields

$$
\begin{aligned}
\tilde{p}_t^{*} = & \int_J \bar{a}\left(1-2^{-2\kappa_i^{(q)*}}\right)\tilde{q}_t dj \\
& -\int_J \left[\sum_{l\neq i}\bar{b}_l\mu_l\left(1-2^{-2\kappa_i^{(l)*}}\right)\tilde{\phi}_{lt}\right] dj \\
& -\int_J \left(\bar{b}_i\mu_i+\gamma\right)\left(1-2^{-2\kappa_i^{(i)*}}\right)\tilde{\phi}_{it} dj
\end{aligned}
$$

notice that the noise variables are mean zero and firm-specific so that the integral with respect to a continuum of firms is equal to zero. Uhlig (1996)

Then we have

$$
\begin{aligned}
\tilde{p}_t^* = & \left[\sum_{i=1}^{I} \mu_i \bar{a} \left(1 - 2^{-2\kappa_i^{(q)*}}\right)\right] \tilde{q}_t - \sum_{i=1}^{I} \mu_i \left[\sum_{l \neq i} \bar{b}_l \mu_l \left(1 - 2^{-2\kappa_i^{(l)*}}\right) \tilde{\phi}_{lt}\right] \\
& - \sum_{i=1}^{I} \mu_i (\bar{b}_i \mu_i + \gamma) \left(1 - 2^{-2\kappa_i^{(i)*}}\right) \tilde{\phi}_{it} \\
= & \left[\sum_{i=1}^{I} \mu_i \bar{a} \left(1 - 2^{-2\kappa_i^{(q)*}}\right)\right] \tilde{q}_t \\
& - \frac{\gamma}{\zeta} \sum_{i=1}^{I} \mu_i \left[\sum_{l=1}^{I} (1-\zeta) b_l \mu_l \left(1 - 2^{-2\kappa_i^{(l)*}}\right) \tilde{\phi}_{lt}\right] \\
& - \frac{\gamma}{\zeta} \sum_{i=1}^{I} \zeta \mu_i \left(1 - 2^{-2\kappa_i^{(i)*}}\right) \tilde{\phi}_{it} \\
= & \left[\sum_{i=1}^{I} \mu_i \left[(1-\zeta) a + \zeta\right] \left(1 - 2^{-2\kappa_i^{(q)*}}\right)\right] \tilde{q}_t \\
& - \frac{\gamma}{\zeta} \sum_{l=1}^{I} \left[(1-\zeta) b_l \sum_{i=1}^{I} \mu_i \left(1 - 2^{-2\kappa_i^{(l)*}}\right)\right] \mu_l \tilde{\phi}_{lt} \\
& - \frac{\gamma}{\zeta} \sum_{l=1}^{I} \zeta \left(1 - 2^{-2\kappa_l^{(l)*}}\right) \mu_l \tilde{\phi}_{lt} \\
= & \left[\sum_{i=1}^{I} \mu_i \left[(1-\zeta) a + \zeta\right] \left(1 - 2^{-2\kappa_i^{(q)*}}\right)\right] \tilde{q}_t - \frac{\gamma}{\zeta} \sum_{l=1}^{I} \left[\sum_{i=1}^{I} w_{li} \left(1 - 2^{-2\kappa_i^{(l)*}}\right)\right] \mu_l \tilde{\phi}_{lt}
\end{aligned}
$$

where $w_{il} = (1-\zeta) b_l \mu_i + \zeta \mathbf{1}(l = i)$ and $\mathbf{1}(l = i)$ is the indicator function that takes the value 1 if $l = i$ and is 0 otherwise. This verifies the guess with

$$
\begin{aligned}
a &= \left[\sum_{i=1}^{I} \mu_i \left[(1-\zeta) a + \zeta\right] \left(1 - 2^{-2\kappa_i^{(q)*}}\right)\right] \\
b_l &= \left[\sum_{i=1}^{I} \left[(1-\zeta) b_l \mu_i + \zeta \mathbf{1}(l = i)\right] \left(1 - 2^{-2\kappa_i^{(l)*}}\right)\right]
\end{aligned}
$$

these can be rewritten to emphasize the effect of the parameter of strategic complementarities

$$a = (1-\zeta)a\sum_{i=1}^{I}\mu_i\left(1-2^{-2\kappa_i^{(q)^*}}\right)+\zeta\sum_{l=1}^{I}\mu_l\left(1-2^{-2\kappa_l^{(q)^*}}\right)$$
$$b_l = (1-\zeta)b_l\sum_{i=1}^{I}\mu_i\left(1-2^{-2\kappa_i^{(l)^*}}\right)+\zeta\left(1-2^{-2\kappa_l^{(l)^*}}\right)$$

Notation

Parameters

Indices

- $h \in H$ index for households
- μ_H measure for households with $\mu_H(H) = 1$
- $j \in J$ index for firms
- μ_J measure for households with $\mu_J(J) = 1$
- $\mu_i \equiv \mu_J(J_i)$ convenience notation
- $i \in \{1, \ldots, I\}$ index for sectors
- $\{J_1, \ldots, J_I\}$ partition on firms induced by sectors

Households

- σ: coefficient of relative risk aversion; inverse of the elasticity of intertemporal substitution parameter

- ε: inverse of Frisch elasticity of labor supply "measures the substitution effect of a change in the wage rate on labor supply." Comes from the derived Household optimization equation $w_t = c_t^{\sigma} n_t^{\varepsilon}$, so that $n_t = w_t^{\frac{1}{\varepsilon}} c_t^{\frac{\sigma}{\varepsilon}}$.

- $r \in [0, 1)$: within-industry generalized mean exponent

- $p \in [0, 1)$: between-industry generalized mean exponent

- $\eta = \frac{1}{1-r} \in [1, \infty)$: within-industry elasticity of substitution; measure of market power

- $\rho = \frac{1}{1-p} \in [1, \infty)$: between-industry elasticity of substitution; measure of trade linkages

<u>Equilibrium</u>

- $\alpha = \frac{1}{1+\rho\varepsilon}$ parameterizes strategic complementarities specifically arising from heterogeneous information, see Angeletos and La'O (2010).

- $\gamma = \alpha(1 + \varepsilon)$

- $\zeta = \alpha(\sigma + \varepsilon)$ is the typical New Keynesian parameter governing strategic complementarities generally (see Morris and Shin, 2002, Woodford (2003) chapter 3, and Mankiw and Reis (2010)) and also related to the degree of "real rigidities" (see Ball and Romer, 1990).

Stochastic Processes

- $\Omega = \{\{q_t\}, \{\phi_{1t}\}, \cdots, \{\phi_{It}\}\}$ is an ordered tuple gathering all stochastic processes and indexed by ω.

- ω_l is the l^{th} item in Ω; for example $\omega_1 \equiv q$.

Fundamentals

– $q_t \overset{iid}{\sim} N(0, \sigma_q^2)$: nominal aggregate demand

– $\phi_{it} \overset{iid}{\sim} N(0, \sigma_{\phi_i}^2)$ idiosyncratic productivity shocks

Signals

– $s_{jt}^{(q)} = \tilde{q}_t + \psi_{jt}^{(q)}$ is the signal to firm j related to aggregate conditions. $s_{jt}^{(q)} \sim N(\tilde{q}_t, \sigma_q^2 + \sigma_{\psi_j^{(q)}}^2)$

– $s_{jt}^{(l)} = \tilde{\phi}_{;t} + \psi_{jt}^{(l)}$ is the signal to firm j related to the productivity shock to industry l. $s_{jt}^{(l)} \sim N(\tilde{\phi}_{lt}, \sigma_{\phi_l}^2 + \sigma_{\psi_j^{(l)}}^2)$

Information

– κ represents the Shannon capacity of a channel, measured in bits. This term is also used for the specific parameter describing total capacity available to agents.

– $\kappa_j^{(\omega)} \equiv \mathcal{I}(\omega, s_j^{(\omega)})$ represents the information capacity allocated by firm j to stochastic process ω

– $\kappa_j^{(\omega)^*}$ represents the optimal allocated capacity by firm j

– $\kappa_i^{(\omega)^*}$ represents the optimal capacity by any firm in sector i allocated to stochastic process ω; requires appealing to symmetry

APPENDIX B

APPENDIX TO CHAPTER 3

Factor selection

The baseline model considered in the paper imposes 5 factors as in the related literature. Here we show that this choice does not influence our results. As suggested by Jungbacker and Koopman (2014), it is difficult to interpret the (hundreds of) factor loading coefficients directly, but one way to assess the contributions of each factor is to regress each series separately on each factor and a constant, and the results displayed as an index plot. Fig. 10 shows this plot for the five factors estimated via principal components.

FIGURE 10. Portion of variance of each observed series explained by each factor.

This plot suggests that the vast majority of the common component for the PCE price series is driven only by the first factor. Since the ordering of the factors is not arbitrary - the first factor is the one which explains the most variation in the data - the common component for inflation series will be little affected, regardless of the number of factors chosen.

Another way to assess the consequences of our assumptions is to reconsider the model to extract factors only from the price series (ignoring the macroeconomic indicators, quantity series, and PPI). Fig. 11 shows the contributions of a factor estimated in this way (i.e. only from price series) to all of the series in the model. Its contributions appears almost identical to the first original factor.

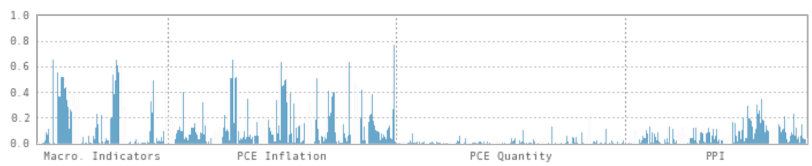

FIGURE 11. Portion of variance of each observed series explained by a single prices-only factor.

To directly check that the single factor extracted from only price series spans the same underlying space as the first factor extracted from all series, we turn to regression analysis. We regress each of the five original factors on the single price factor. The R^2 value from each of these regressions is reported in Table 14.

TABLE 14.
R^2 statistic expressing factor contributions

	R^2
Factor 1	0.953913
Factor 2	0.001450
Factor 3	0.007194
Factor 4	0.001825
Factor 5	0.005854

REFERENCES CITED

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, 80(5):1977–2016.

Alvarez, R., Camacho, M., and Perez-Quiros, G. (2016). Aggregate versus disaggregate information in dynamic factor models. *International Journal of Forecasting*, 32(3):680–694.

Angeletos, G.-M. and La'O, J. (2010). Noisy Business Cycles. *NBER Macroeconomics Annual 2009*, 24:319–378.

Ansley, C. F. and Kohn, R. (1986). A note on reparameterizing a vector autoregressive moving average model to enforce stationarity. *Journal of Statistical Computation and Simulation*, 24(2):99–106.

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics*, 40(1):436–465.

Bai, J. and Li, K. (2015). Maximum Likelihood Estimation and Inference for Approximate Factor Models of High Dimension. *Review of Economics and Statistics*, 98(2):298–309.

Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2006). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2008a). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

Bai, J. and Ng, S. (2008b). Large Dimensional Factor Analysis. *Foundations and Trends(R) in Econometrics*, 3(2):89–163.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29.

Ball, L. and Romer, D. (1990). Real Rigidities and the Non-Neutrality of Money. *The Review of Economic Studies*, 57(2):183–203.

Basu, S. (1995). Intermediate Goods and Business Cycles: Implications for Productivity and Welfare. *The American Economic Review*, 85(3):512–531.

Babura, M. and Modugno, M. (2014). Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data. *Journal of Applied Econometrics*, 29(1):133–160.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics*, 120(1):387–422.

Bils, M. and Klenow, P. (2004). Some Evidence on the Importance of Sticky Prices. *Journal of Political Economy*, 112(5):947–985.

Boivin, J., Giannoni, M. P., and Mihov, I. (2009). Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data. *The American Economic Review*, 99(1):350–384.

Boivin, J. and Ng, S. (2005). Understanding and Comparing Factor-Based Forecasts. Working Paper 11285, National Bureau of Economic Research.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Bouakez, H., Cardia, E., and Ruge-Murcia, F. J. (2009). Sectoral Price Rigidity and Aggregate Dynamics. SSRN Scholarly Paper ID 1548020, Social Science Research Network, Rochester, NY.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.

Carter, C. K. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81(3):541–553.

Carvalho, C. V. d. and Lee, J. W. (2011). Sectoral Price Facts in a Sticky-Price Model. Departmental Working Paper, Rutgers University, Department of Economics.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica*, 51(5):1281–1304.

Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2000). Sticky Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem? *Econometrica*, 68(5):1151–1179.

Choi, I. (2012). EFFICIENT ESTIMATION OF FACTOR MODELS. *Econometric Theory*, 28(02):274–308.

Clements, M. P. and Hendry, D. F. (2011). *The Oxford Handbook of Economic Forecasting*. Oxford University Press.

Cogley, T., Primiceri, G. E., and Sargent, T. J. (2010). Inflation-Gap Persistence in the US. *American Economic Journal: Macroeconomics*, 2(1):43–69.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. John Wiley & Sons.

De Graeve, F. and Walentin, K. (2014). Refining Stylized Facts from Factor Models of Inflation. *Journal of Applied Econometrics*, pages n/a–n/a.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, 67(3):297–308.

Dotsey, M., King, R. G., and Wolman, A. L. (1999). State-Dependent Pricing and the General Equilibrium Dynamics of Money and Output. *The Quarterly Journal of Economics*, 114(2):655–690.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Doz, C., Giannone, D., and Reichlin, L. (2012). A QuasiMaximum Likelihood Approach for Large, Approximate Dynamic Factor Models. *Review of Economics and Statistics*, 94(4):1014–1024.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford University Press.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The Generalized Dynamic Factor Model. *Journal of the American Statistical Association*, 100(471):830–840.

Friedman, M. (1968). The Role of Monetary Policy. *The American Economic Review*, 58(1):1–17.

Geweke, J. (1976). *The dynamic factor analysis of economic time series models*. University of Wisconsin.

Golosov, M. and Lucas Jr., R. E. (2007). Menu Costs and Phillips Curves. *Journal of Political Economy*, 115(2):171–199.

Harvey, A. and Shephard, N. (2005). Structural Time Series Models. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd.

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A. C. and Jaeger, A. (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics*, 8(3):231–247.

Jungbacker, B. and Koopman, S. J. (2008). Likelihood-based Analysis for Dynamic Factor Models. Tinbergen Institute Discussion Paper 08-007/4, Tinbergen Institute.

Jungbacker, B. and Koopman, S. J. (2014). Likelihood-based dynamic factor analysis for measurement and forecasting. *The Econometrics Journal*.

Jungbacker, B., Koopman, S. J., and van der Wel, M. (2011). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control*, 35(8):1358–1368.

Kaufmann, D. and Lein, S. M. (2013). Sticky prices or rational inattention What can we learn from sectoral price data? *European Economic Review*, 64(C):384–394.

Klenow, P. J. and Kryvtsov, O. (2008). State-Dependent or Time-Dependent Pricing: Does it Matter for Recent U.S. Inflation? *The Quarterly Journal of Economics*, 123(3):863–904.

Koehler, A. B. and Murphree, E. S. (1988). A Comparison of the Akaike and Schwarz Criteria for Selecting Model Order. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 37(2):187–195.

Koop, G. (2003). *Bayesian Econometrics*. Wiley-Interscience, Chichester ; Hoboken, N.J, 1st edition.

Koopman, S. J. and Durbin, J. (2000). Fast Filtering and Smoothing for Multivariate State Space Models. *Journal of Time Series Analysis*, 21(3):281–296.

Koopman, S. J. and Shephard, N. (1992). Exact Score for Time Series Models in State Space Form. *Biometrika*, 79(4):823–826.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2003). International Business Cycles: World, Region, and Country-Specific Factors. *The American Economic Review*, 93(4):1216–1239.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, 4(2):103–124.

Mankiw, N. G. and Reis, R. (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics*, 117(4):1295–1328.

Mankiw, N. G. and Reis, R. (2010). Imperfect Information and Aggregate Supply. Handbook of Monetary Economics, Elsevier.

Makowiak, B., Moench, E., and Wiederholt, M. (2009). Sectoral price data and models of price setting. *Journal of Monetary Economics*, 56, Supplement:S78–S99.

Makowiak, B. and Wiederholt, M. (2009). Optimal Sticky Prices under Rational Inattention. *The American Economic Review*, 99(3):769–803.

Morris, S. and Shin, H. S. (2002). Social Value of Public Information. *The American Economic Review*, 92(5):1521–1534.

Mller, U. K. (2013). Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix. *Econometrica*, 81(5):1805–1849.

Nakamura, E. and Steinsson, J. (2008). Five Facts about Prices: A Reevaluation of Menu Cost Models. *The Quarterly Journal of Economics*, 123(4):1415–1464.

Nakamura, E. and Steinsson, J. (2010). Monetary Non-neutrality in a Multisector Menu Cost Model. *The Quarterly Journal of Economics*, 125(3):961–1013.

Otrok, C. and Whiteman, C. H. (1998). Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa. *International Economic Review*, 39(4):997–1014.

Phelps, E. S. (1968). Money-Wage Dynamics and Labor-Market Equilibrium. *Journal of Political Economy*, 76.

Reis, R. and Watson, M. W. (2010). Relative Goods' Prices, Pure Inflation, and The Phillips Correlation. *American Economic Journal: Macroeconomics*, 2(3):128–157.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. Working Paper 55, Federal Reserve Bank of Minneapolis.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423.

Sims, C. (2010). Rational Inattention and Monetary Economics. Handbook of Monetary Economics, Elsevier.

Sims, C. A. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy*, 49:317–356.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Sims, C. A. (2005). Rational inattention: a research agenda. Discussion Paper Series 1: Economic Studies 2005,34, Deutsche Bundesbank, Research Centre.

Sneek, J. M. (1984). *Modelling Procedures for Univariate Economic Time Series.* Vu Uitgeverij, Amsterdam.

Stock, J. H. and Watson, M. W. (2002a). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, 39:3–33.

Stock, J. H. and Watson, M. W. (2011). Dynamic factor models. In *Oxford Handbook of Economic Forecasting.*

Townsend, R. M. (1983). Forecasting the Forecasts of Others. *Journal of Political Economy*, 91(4):546–588.

Uhlig, H. (1996). A Law of Large Numbers for Large Economies. *Economic Theory*, 8(1):41–50.

Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400.

West, M. and Harrison, J. (1999). *Bayesian Forecasting and Dynamic Models.* Springer, New York, 2nd edition.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50(1):1–25.

Woodford, M. (2001). Imperfect Common Knowledge and the Effects of Monetary Policy. Working Paper 8673, National Bureau of Economic Research.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.